ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

Application of SVM for cell recognition in BCC skin pathology*

Tomasz Markiewicz[1,2], Stanislaw Osowski[1,3], Cezary Jochymski[2], Joanna Narbutt[4], Wojciech Kozlowski[2]

1- Warsaw University of Technology – Dept of Electrical Engineering
ul. Koszykowa 75, 00-661 Warsaw - Poland, tel. +48222347235, markiewt@iem.pw.edu.pl
2- Military Institute of the Health Services - Dept of Pathology
ul. Szaserow 128, 00-909 Warsaw - Poland
3- Military University of Technology - Dept of Electronics
ul. Kaliskiego 2, 00-908 Warsaw - Poland
4- Medical University of Lodz - Dept of Dermatology and Venerology
ul. Krzemieniecka 5, 94-017 Lodz - Poland

Abstract.

The paper presents the application of Support Vector Machine (SVM) for the recognition of immunopositive and immunonegative cells at basal cell carcinoma. The developed algorithm applies two kinds of SVM: the Gaussian kernel SVM for direct cell recognition and linear kernel SVM as a preprocessing stage for sequential thresholding of the image. The developed computer program was tested on the examples of 528 images of carcinoma and the obtained results are in good agreement with the human expert score.

1 Introduction

The keratinocytic tumors are a clinically and histopathology diverse group of lesions derived from the proliferation of epidermal and adnexal keratinocytes. These tumors account for approximately 90% or more of all skin malignancies, of which about 70% are basal cell carcinoma (BCC). BCC is a group of malignant cutaneous tumors characterized by the presence of lobules, columns, bands or cords of basaloid cells [1]. The multiple variants of BCC are characterized by common histological feature.

The key point in the medical investigations of the BCC is the recognition of the immunonegative and immunopositive cells after application of some set of the antibodies to the biopsy of skin. The research in the field of carcinoma is concentrated on the quantitative evaluation and based on the recognition and counting the quantity of the immunopositive and immunonegative cells. There is no such automatic system for BCC on the market at the moment [1,2,3]. The main task of this work is to develop the computer program able for the automatic extraction and recognition of both types of the immunoreactive cells on the basis of BCC image.

* This work is supported by Polish Ministry of Science and Higher Education by grant in the years 2006-2009.

2 Data base

The specimens used in experiments have been stained for 11 primary antibodies: p16, p18, p21, p27, p53, p63, Cyclin A, Cyclin B1, Cyclin D1, Bax and BCL2. The evaluated areas of the image have been selected by the medical human expert and contain the regions of skin full of neoplasm cells and also healthy epidermis area. The skin specimens of 12 patients with nodular type and 12 patients with superficial type of BCC have taken part in investigations. For all investigated patients the mentioned above 11 stains were applied to the healthy skin and to the neoplasm. In this way we have investigated 528 images in the aspect of the immunoreactivity of cells.

3 The applied methods of image processing

The research of the BCC skin pathology is directed to the recognition of the stained cells with the specific antibodies. Especially important is the determination of the distribution density of these cells and this task needs the correct recognition of the separated nuclei and their immunoreactivity.

All applied antibodies stain nuclei structures of cells according to their immunoreactivities. At the correct staining the nuclei of the cells with negative reaction to the antibody (the immunonegative cells) are colored in blue and the cells of positive reaction to antibody (the immunopositive cells) are colored in brown. The problem of recognition between these two types of cells requires solved the two tasks: the extraction of the nuclei of the cells from the whole image and classification of them into two groups, based on the color of their nuclei.

The usual form of the input image of the skin sample is in the form of RGB representation. The first step of the processing is the image standardization. It is done by using also other (NTSC) representation of the image. The standardized form of the image is formed on the pixel basis and is defined in the following form

$$f(x,y,s)_{st} = \frac{f(x,y,s) - NTSC(1)_{\min}}{NTSC(1)_{\max} - NTSC(1)_{\min}} \tag{1}$$

where s represents each of RGB component, x and y are the pixel positions, $f(x,y,s)$ is the original intensity of sth color component of the pixel at (x,y) position. The $NTSC(1)$ represents the first $NTSC$ component (luminance) of the image, while $NTSC_{min}$ and $NTSC_{max}$ are the minimum and maximum value of this component.

The standardization described by equation (1) reduces the influence of the differences in the glass transparency, non-uniform distribution of staining, effects of non-equal lighting and other imperfections at the image production stage. It is very significant step in image preprocessing because the cells in original image are stained non-uniformly with the significant variation of blue (brown) intensities of cells. The standardized image contains more compact and uniform colors of both types of cells.

The next step of image processing is the extraction of cells on the basis of recognition of their nuclei. The most often used approach applies the threshold operation, defined in the form [4]

$$T_{[0,t]}[f(x,y)] = \begin{cases} 1 & \text{if} \quad 0 \le f(x,y) \le t \\ 0 & \text{else} \end{cases} \tag{2}$$

where $f(x,y)$ is the value of pixel intensity of the standardized image f at the position (x,y) and t is the threshold value. This operation is usually done on the grayscale image representation, obtained from its color version. The important problem in this method is the choice of the threshold value, since the cells are non-uniformly stained. Hence application of even optimally chosen threshold (for example by applying Otsu method [4]) results in some imperfections and omission of some cells.

In our solution to this problem we propose the extraction algorithm based on the application of Support Vector Machine (SVM) and sequential thresholding. The first step of this approach is to recognize the blue and brown cells from the lighter background. Each of them will represent the appropriate class. In this way we define 3 classes (the first – blue, the second – brown and the third – background). The classification task was solved using the SVM network of Gaussian kernel working in the classification mode [5]. The hyperparameters (the regularization constant and width of the Gaussian function) have been selected using the crossvalidation technique on the additional 4 validation images. The input vector $\mathbf{x}$ for SVM was formed from 3 RGB components of the standardized image pixels. The learning data were manually selected from the appropriate regions of 4 images chosen for learning purposes. They should be characteristic for the blue, brown and background fields. These data formed the learning set for SVM classifiers. On the experimental way we decided to use 150 pixels for each class.

Because one SVM network recognizes only two classes, we have solved this 3 class recognition problem using three SVM networks working in one-against-one mode [5]. As a result of this step we have split the whole image into the regions belonging to three classes: the blue pixel class, brown pixel class and the background. For further procedure only two classes are relevant: the blue pixels class and brown pixel class. They will form the introductory masks of blue and brown cells.

In the second parallel step of cell extraction and recognition we propose the application of the thresholding operation applied sequentially at different values of the threshold. However to get the best possible results we perform the thresholding not on the original image but on its transformed representation. We form two transformed images. The first one corresponds to the blue cells and background, and the second image corresponds to the brown cells and the background (these two kinds of data are used at training of each SVM). The data for these two classes come from the same standardized image and are in identical form as for Gaussian kernel SVM classifier.

This time we apply two linear kernel SVM networks: one for the recognition between the blue cells (destination +1) and the background (destination -1) and the second for the recognition between the brown cells (destination +1) and the background (destination -1). The output signal of linear kernel SVM is described as

$$D(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b_0 \tag{3}$$

where $\mathbf{w}$ is the weight vector and b is the bias. The input vector for both SVM networks is composed of three color components of the pixel in RGB standard.

In the testing phase (after fixing the learned SVM structure and parameters) the input vectors for these 2 SVM networks represent the same 3 color intensities of all pixels of the images not taking part in learning. It is evident that the output value $D(\mathbf{x})$ corresponding to the class of blue pixels (first SVM) will be different, depending on their actual relation to the background. The same is true for the second SVM (brown pixels). For example, if the actual pixel is light blue, the $D(\mathbf{x})$ signal of the SVM

network will take values $0<D(\mathbf{x})<1$. For the dark blue pixels this value will be higher than 1. Based on this finding we use $D(\mathbf{x})$ value as an indicator of the intensities of blue and brown pixels with reference to the background. In this way the original image was converted to two other images: one presenting $D(\mathbf{x})$ values for blue-background data and the second of brown-background data.

These two transformed images are subject to the sequential thresholding operations starting from the minimum $D(\mathbf{x})$ used as a threshold value. This value is increased step by step until its maximum. In any step of thresholding the certain image area is separated and tested as the candidate for the cell. First it is checked if the separated area is not already found by Gaussian kernel SVM classifier used in the first step. If not, we check if the area of the compact separated region fulfills the strictly specified size limit of the possible cells (the minimum 50 pixels, the maximum 150 pixels). If it falls within these limits the separated objects are added to the already existing masks of the blue or brown cells. As a result of this we get the masks of all blue and brown cells, satisfying the preselected range of their size.

However there is some problem with determination of the maximum size of the cells in the thresholding operation. It is due to the fact that some cells may be glued together and form bigger objects. To get the highest possible accuracy of the extraction algorithm we have to split such cells. We have done it by applying the watershed operation [3] applied only for the cells exceeding the size of 150 pixels. This operation divides such objects into few cells if there is narrow lighter space between two or more parts of the object. In this way we are able to get better accuracy of cell recognition of the image. In practice we have noticed that this stage of algorithm was able to discover around 10% of total population of cells in the image.

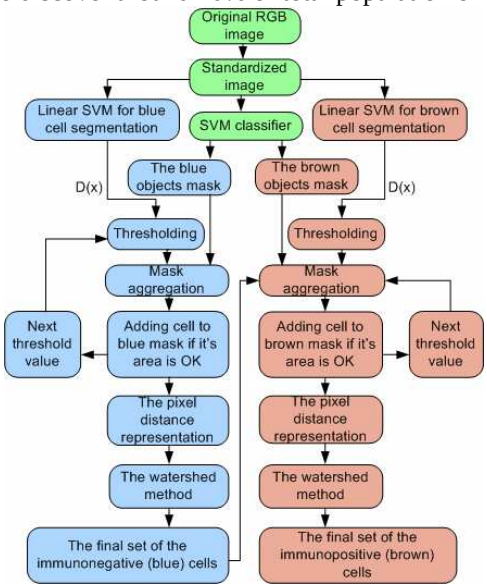

Fig. 1: The diagram of the automatic system for cell recognition

The final task of image processing is to count the extracted cells belonging to both classes of cells (brown cells – the immunopositive class, blue cells – the

immunonegative class). This is simple task in Matlab platform implementation [6], since we got two different masks of cells, separately brown and separately blue cells.

Fig. 1 presents the block diagram of the presented above algorithm. There are two parallel ways of image processing: the SVM Gaussian kernel classifier system recognizing directly blue and brown objects, and the sequential thresholding operations supported by the linear kernel SVM. In both stages the lower and upper limits of the cell size are checked. As a result we get the final quantity of both types of cells in the selected view field of the image. The algorithm was implemented in the form of automatic program written on the Matlab platform with GUI interface.

4 Results

The designed automatic system of cell recognition was tested on a wide basis of many images (528 images not taking part in learning and validation). Its main application is the quantitative evaluation of the skin specimens stained by the eleven specified antibodies. In its practical implementation the human expert is responsible only for the selection of the neoplasm area and for the healthy epidermis area of the skin. The acquired images are evaluated by the computer system automatically. Each analyzed image has been also annotated automatically by the system. The immunopositive cells are marked by the sign of "○" of the red color and the immunonegative cells by "∇" of black colors respectively.

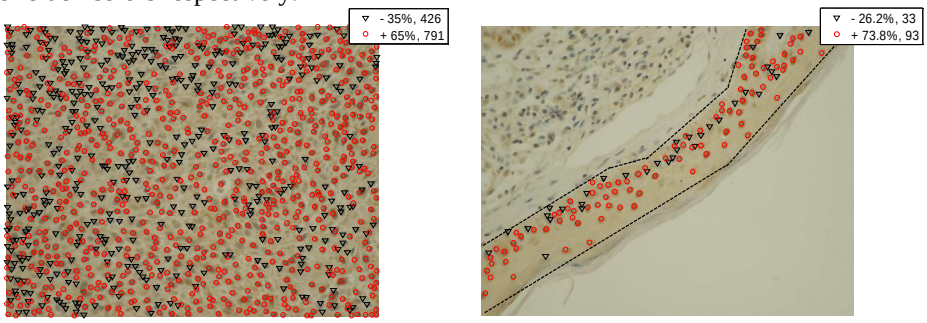

Fig 2 The sample results of the recognition and counting of the immunopositive and immunonegative cells in a) neoplasm area (Bax) and b) healthy epidermis area (p63, 400x)

Figure 2 presents the graphical sample results of the cell recognition at the Bax stain of the neoplasm (Fig. 2a) and healthy epidermis area (Fig. 2b) selected by the expert (the area within dashed line). The important factor in BCC research is the ratio of the immunonegative and immunopositive cells for both regions (the neoplasm and healthy epidermis areas)

To get the reliable results of testing we have performed the analysis of 528 images corresponding to 24 patients at application of 11 staining antibodies and 2 places of image acquisition (the neoplasm and healthy epidermis areas). The results have been compared to the score of the human expert. We have got very good agreement with the human score. Table 1 presents the quantitative results of the immunoreactive (IN - immunonegative and IP – the immunopositive) cells for 6 chosen images at Bax staining. The symbol AS corresponds to the results of our

automatic system and EXP to the human expert. The overall discrepancy between the AS and EXP scores for these 528 images was equal 4.7% and this accuracy is fully acceptable in medical practice (10% of difference of results are acceptable).

Patient	AS			EXP		
	No of IN cells	No of IP cells	$\frac{IP}{IP+IN}$	No of IN cells	No of IP cells	$\frac{IP}{IP+IN}$
1 (neoplasm)	426	791	65.0%	439	784	64.1%
1 (healthy)	18	108	85.7%	19	100	84.3%
2 (neoplasm)	1131	413	26.8%	1116	428	27.7%
2 (healthy)	11	78	87.6%	10	75	88.2%
3 (neoplasm)	107	624	85.4%	119	609	83.4%
3 (healthy)	11	144	92.9%	12	146	92.4%

Table 1 The detailed results of the recognition of the immunoreactive cells for 6 chosen images of the BCC at Bax staining for 3 different patients

The main advantage of the developed program is its speed and reliability. At its Matlab implementation on PC Centrino Duo of 1.86 GHz, 2GB RAM we got the final score for one image in less 1 minute. For human expert it takes around 10 minutes. Moreover the results of our program are repeatable. This is not true for the human experts. Even the same expert may produce different results for the same image. The actual results depend on his physiological state, fatigue and actual concentration.

Additionally we have developed the graphical tools for the human expert to intervene into the final recognition results. If, according to him some cells are wrongly classified he can change their class membership in a manual way. At the same time his correction is added automatically to the numerical results concerning the number of immunopositive and immunonegative cells.

5 Conclusions

The paper has presented the automatic computerized system for recognition of the immunopositive and immunonegative cells of the skin and skin neoplasm biopsies. The main role in this system fulfills the Support Vector Machine, responsible for the extraction of both types of cells. The system was checked on a wide basis of images corresponding to 528 images acquired from 24 patients suffering from BCC.

References

[1] S. Kossard, E.H. Epstein Jr, R. Cerio, L.L. Yu, D. Weedon, Basal cell carcinoma In D. Weedon, P. LeBoit, G. Burg, A. Sarasin (editors), Pathology and Genetic of Tumors of the Skin, International Agency for Research on Cancer (IARC), pages 13-19, Lyon, 2000

[2] G. Kayser, D. Radziszowski, P. Bzdyl, R. Sommer, K. Kayser, Theory and implementation of an electronic, automated measurement system for images obtained from immunohistochemically stained slides, *Anal Quant Cytol Histol*, 28:27- 38, SPP, Inc., 2006.

[3] T. Markiewicz, S. Osowski, J. Patera, W. Kozlowski, Image processing for accurate recognition and counting of cells of the histological slides, *Anal Quant Cytol Histol*, 26:281:292, SPP, Inc., 2006.

[4] P. Soille: Morphological image analysis, principles and application, Springer, Berlin, 2003.

[5] B. Schölkopf, A. Smola: Learning with Kernels, MIT Press, Cambridge, MA, 2002

[6] Matlab Image Processing Toolbox, user's guide, MathWorks, Natick, 2002