AUTOMATIC MEMBRANE SEGMENTATION OF IHC IMAGES ENABLES THE ANALYSIS OF CANCER BREAST TISSUES

Raquel Pezoa$^{1,2}$, Rodrigo Rojas-Moraleda$^{1,2}$ and Luis Salinas$^{1,2}$

$^1$ Informatics Department, Universidad Técnica Federico Santa María, Valparaíso, Chile
$^2$ Centro Científico Tecnológico de Valparaíso, Universidad Técnica Federico Santa María, Valparaíso, Chile
raquel.pezoa@usm.cl, rodrigo.rojas@postgrado.usm.cl, luis.salinas@usm.cl

Keywords: histological image processing, support vector machines, immunohistochemistry

1 INTRODUCTION

The automatic quantification in histological studies plays an important role in Histopathology and basic research. Computational methods capable of providing accurate and objective results are fundamental in several medical applications such as automatic membrane and nuclei quantification in breast cancer tissues.

The basis for a robust and accurate quantification is the segmentation of regions of interest (ROIs) defined by different histopathologic features. Several kinds of problems arise in ROIs segmentation of histopathologic images due to diverse factors such as: nonstationary and correlated noise, illumination, busyness of gray levels within the object and its background, inadequate contrast, among others [7].

This study is focused on segmentation and quantification of structural elements of histopathology images, in particular, of immunohistochemistry stained (IHC-stained) breast cancer tissue images. The assessment of HER2 (human epidermal growth factor receptor 2) protein overexpression is fundamental for the prognosis of breast cancer and for patient treatment as well [1, 4]. When IHC is applied to the tissue, HER2 overexpression is displayed by brown staining of membrane elements (see Fig. 1). As a rule, pathologists determine HER2 overexpression through visual inspection and estimate two parameters: intensity and completeness. The former parameter corresponds to the intensity of brown staining and the latter corresponds to the percentage of staining completeness with respect to the membrane length. According to the two estimated parameters the observed specimen is classified [9] into four categories: 0, 1+, 2+ and 3+. Here 0 means that no staining was observed, 1+, 2+ and 3+ mean, respectively, that barely perceptible, weak to moderate and strong intensity was observed in more than 10% of tumor cells.
The approach in the present paper is mainly based on the use of SVM for the segmentation of membranes elements. SVM is a well known binary classifier [5], and for this paper it has been used to classify the image pixels into two classes: overexpressed and non-overexpressed. For nuclei segmentation, the Otsu’s thresholding method [6] was used.

Once the ROIs (membranes and nuclei) are identified, it is necessary to reconstruct the cancer cell membrane in order to obtain the intensity and completeness of the HER2 overexpression, for which an approach of minimum distance criterion from nuclear boundaries was used.

2 METHODS

This section describes the sequence of steps undertaken in order to analyze the IHC-stained images. Details about the image acquisition and image segmentation workflow are described as follows.

2.1 Image Acquisition

26 glass slides were automatically imaged in bright-field mode at a 40-fold magnification (0.23 µm/pixel) using the Hamamatsu NanoZoomer 2.0-HT Scan System (Hamamatsu Photonics, Japan. The resulting virtual slides (see Fig. 2) have an average file size of 765 MB in JPEG compression format (and around 15 GB when is uncompressed).
Three predetermined squared regions of interest with spatial dimension 133 µm x 133 µm were extracted (a total of 78 IHC-stained images) by a pathologist from each virtual slide in order to develop the segmentation of membrane and nuclei cell elements and the HER2 overexpression quantification.

2.2 Segmentation Workflow

The full segmentation workflow is achieved by two main stages: (i) SVM training and (ii) membrane and nuclei segmentation.

2.2.1 SVM Training

Fig. 3 shows the main three steps of SVM training stage:

Feature extraction is an important consideration in SVM method and finding an appropriate space feature is a complex task. In this study 11 features were analyzed and combined, such as standard deviation, entropy, dynamic range, Sobel gradient magnitude, Y (from CMYK) channel pixel, Jensen-Shannon divergence, among others [3].

78 manually drawn masks were generated by an expert pathologist, who drawn contours that represent the HER2 overexpression. Each of the 78 IHC-stained images contains 335,400 pixels, and SVM training data sets were generated for each image taking pixels belonging to the ROI marked by the pathologist (HER2 overexpression) and pixels belonging to the background (non HER2 overexpression). Each pixel is represented by a vector of 11 features and a binary label indicating if the pixel belongs to the ROI.

The libsvm [2] library was used for SVM training and further classification. The selection of a suitable kernel function is an important problem confronting SVM setup due to a properly selection can minimize the generalization error. In this work the Radial Basis
Function (RBF) was selected. This selection which is based on performance measurements obtained in our previous work [3]. A conventional five-fold cross-validation procedure was performed for SVM training, parameter optimization and testing.

Training was performed independently for each one of the 78 images in order to obtain a ground level for the classification performance. Later, 12 new artificial training sets were obtained by mixing the training information of the 78 stained-images. Thus, 12 models of classifiers were developed and evaluated over the 78 images. The classifier with better generalization on these sets was chosen.

2.2.2 Membrane and Nuclei Segmentation

Fig. 4 shows the main five steps of SVM training stage:

Figure 4. Workflow diagram depicting Membrane and nuclei segmentation steps.

Pixels classification was developed using the SVM classifier model obtained in the previous stage. The result was a binary image (called IHC-svm-membrane image) where 1 means HER2 overexpression and 0 means non-overexpression. Later, the nuclei segmentation was developed using the Otsu’s thresholding method which was applied to the H channel (from HSV color model) of the IHC-stained image.

Automatic HER2 overexpression quantification is particularly hard due to the fact that in immunohistochemical tissues cellular membranes are visible only in the stained tracts of the cell, while the unstained tracts are barely visible. Based in reference [8], a morphological method was developed in order to reconstruct the membrane in the unstained tracts. This method also produced a binary image (called IHC-reconstructed image) where value 1 is given to membrane and 0 to background pixels.

Once the membranes contours are obtained it is possible to measure the completeness parameter for each cell. The average of completeness values correspond to the continuity of the IHC-stained image.

Finally, the intensity parameter was obtained using the following images: IHC-reconstructed image, IHC-svm-membrane image and the Y channel (from CMYK color model) of the original IHC-stained. The intensity of the IHC-image corresponds to the average of the intensities of each cell.
3 RESULTS AND DISCUSSION

Preliminary results have been generated using a set of 13 IHC-stained images. Fig. 5 shows the segmentation of a 2+ IHC-stained image, evaluated with “medium” intensity (from weak, medium and strong) and 85% of completeness by the pathologist. The proposed method gave an intensity of 53% (0% means no brown intensity and 100% means strong brown intensity) and 73% of completeness.

The method gave acceptable results for images belonging to categories 2+ and 3+. Clearly, the lack of an appropriate gold standard for HER2 overexpression quantification impedes a correct validation of the proposed method.

The SVM allows to: (i) face the ambiguity in the segmentation problem, introducing knowledge during the training process (ii) develop a flexible description for membrane/background by means of a multi-parametric feature space of color and texture properties, (iii) create a mechanism to incorporate new sources of information. The morphological step allows imposing constraints over the SVM solutions, based on the nuclei and membrane spatial properties.

4 CONCLUSIONS

The proposed method approach relies on Support Vector Machines (SVM) for membrane segmentation and Otsu’s thresholding for nuclei segmentation. The method offers flexibility and is consistently well suited for two of the most severe HER2 overexpression (2+ and 3+).

Some drawbacks are: (i) this approach is highly dependent of prior knowledge, (ii) SVM introduces the problem of kernel choice, (iii) morphological constraints impose hard assumptions over the cells shape, and (iv) the mixture of morphological and machine learning methods is sensible to noise. In addition, the validation of the algorithm is a fundamental step and expert guided definition of appropriate ground truth data is fundamental. Appropriate gold standard is not available and part of the future work is to develop a system that allows generating ground truth data in a simple way.

ACKNOWLEDGEMENTS

This research is partially funded by CCTVAL-FB0821, Anillo Act 119 and Fondecyt Grant 1100805.

The authors would like to thank Dr. Raúl González and his collaborators at the Gustavo Fricke Hospital, Viña del Mar, Chile, for providing the images and for many helpful discussions as well. Furthermore, we deeply appreciate the support received from
REFERENCES


