Advanced Cancer Cell Characterization and Quantification of Microscopy Images

Theodosios Goudas and Ilias Maglogiannis

University of Central Greece
Department of Computer Science and Biomedical Informatics, Lamia, Greece
{goudas,imaglo}@ucg.gr

Abstract. In this paper we present an advanced image analysis tool for the accurate characterization and quantification of cancer and apoptotic cells in microscopy images. Adaptive thresholding and Support Vector Machines classifiers were utilized for this purpose. The segmentation results are improved through the application of morphological operators such as Majority Voting and a Watershed technique. The proposed tool was evaluated on breast cancer images by medical experts and the results were accurate and reproducible.

Keywords: Image Analysis, SVM, Breast Cancer, Cancer cell, Adaptive Thresholding, Watershed, MCF-7.

1 Introduction

The analysis and characterization of biomedical image data is a complex procedure involving several processing phases, like data acquisition, pre-processing, segmentation, feature extraction and classification. The proper combination and parameterization of the above phases enables the development of adjunct tools that can help on the early diagnosis or the monitoring of therapeutic procedures. A specific imaging domain, which image analysis and processing techniques may apply, is that of histological images obtained by optical or electronic microscopy. For instance Loukas et al [6], achieved cell counting in complex large scale histological images by utilizing edge and color information. Saveliev and Pahwa [7] approached the cell counting problem using a topological working algorithm on binary and grayscale images, suitable for different types of cells. Phukpattaranont and Boonyaphiphat [8] presented an algorithm for the segmentation of cancer cells on microscopic images of immunohistologically stained slides from breast cancer. Their method was based on colour contents, neural networks classification and cell size consideration. Maglogiannis et al [9] successfully classified biological microscopic images of lung tissue sections with idiopathic pulmonary fibrosis. Tosun and Gunduz-Demir [10] proposed an effective on histopathological images segmentation algorithm. Especially the evaluation of cancer related pathology images is considered quite important, since it requires years of theoretical and practical training for a pathologist in order to be able to recognize and diagnose fast and accurately the status and the future evolution of a tumor cells.

I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, pp. 315–322, 2012.

Breast cancer on the other hand is one of the most commonly diagnosed cancer and one of the leading causes of death in United States [1]. Radiotherapy is one of the most common treatments on breast cancer cases [2]. Combined modality treatments are developed through empirical approaches using specific drugs in order to stop the growth of tumors and cause the cancer cell to enter the apoptotic phase with the assistance of the external beam radiation therapy. Researches on Vitamin D [3] proved effectiveness against a broad range of tumor cell types.

The goal of the specific work is the development of an advanced image analysis tool for the accurate quantification of the cancer and the apoptotic cells and the recognition of the corresponding regions in the microscopy image. The tool is used for the evaluation of the influence of the vitamin D3 analogue EB1089 [5] with fractionated radiation on growth and apoptosis human breast cancer tumour MCF-7 [11] cells injected in mice. The rest of the paper is organised as follows; Section II describes the utilized tools and methods for developing the image analysis tool. Section III presents the evaluation, while Section IV concludes the paper.

2 Materials and Methods

Fig. 1 illustrates the block diagram of the proposed tool. Images are first edited by the classification model, which accomplishes the separation of the different type of cells into their respective generated images. The training of the above model is based on the ground truth provided by expert pathologists.

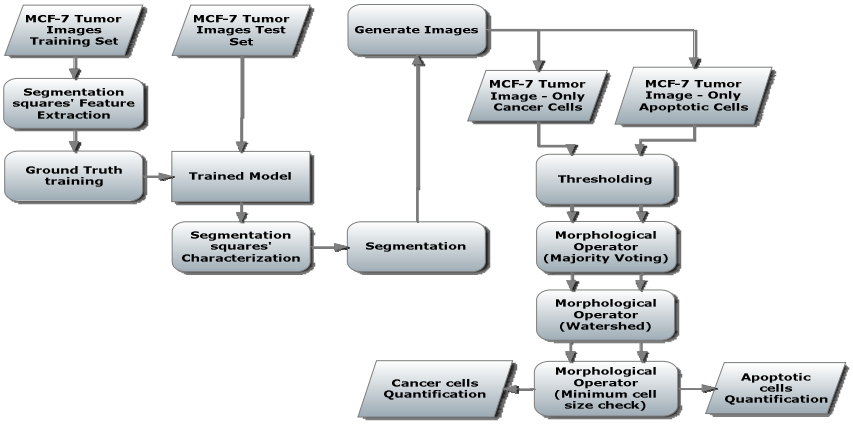

Fig. 1. Data Flow Block Diagram of the Proposed Tool

2.1 Image Dataset Description

The microscopy images dataset was obtained from the National Cancer Institute tumor repository (Frederick, MD) [19]. These images were taken from six week old mices, which were injected the MCF-7 human breast cancer type. Subconfluent cultures, which were grown in RPMI 1640 on 37oC, were fed with fresh medium, washed with PBS,

trypsinized, resuspended in medium, and pooled. After centrifugation, cells were resuspended in Matrigel and cold RPMI 1640 for s.c. inoculation in mice.

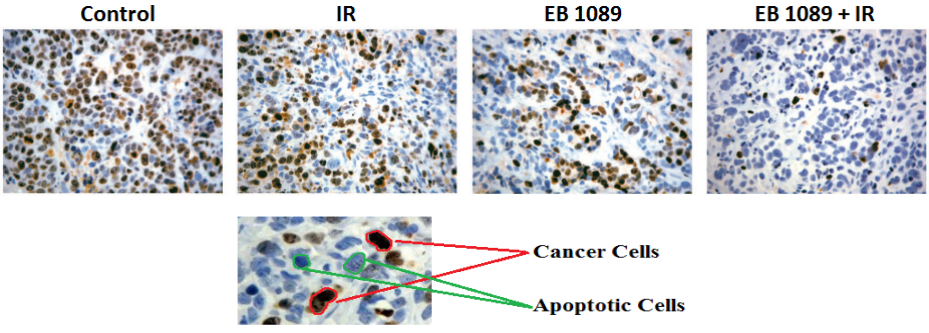

Fig. 2. Four treatment groups of cancer cell images and indication of Cancer and Apoptotic Cells

When the tumor volumes reached 150–200 mm3, tumor-infected mice were randomly selected to receive 3 different treatments. The first group's mice received EB 1089 alone (45 pmol/24 h for 8 days), the second group received radiation alone and the third EB 1089 followed by radiation. So the image dataset provided four groups of datasets, Control, IR (Radiation), EB 1089 (Drug) and EB 1089 + IR combined (see Fig. 2). As also depicted in the second row of Fig. 2, cancer and apoptotic cells have a circular form and sometimes are merged making harder their quantification.

2.2 Feature Extraction and Classification

Customized code was developed, utilizing the ImageJ [12] library, which separates cancer and apoptotic cells from the entire image. While four classifiers were evaluated (Support Vector Machines (SVM), Naïve Bayes, K- Nearest Neighbor and Decision Trees) the SVM classifier was selected since it achieved the higher rates. The mean values of red, green and blue color channels are the features used for the training and testing of the classification model. The ten-fold cross validation has been adopted as a method for testing the accuracy of the classification model. The width of the segmentation square was selected heuristically, from a 2 to 8 pixels range, and set at 2. The case of classifying one by one the pixels was avoided because of the increased processing time during classification and the increased possibility of multiple misclassification issues.

More specifically, the entire image is scanned into segmentation squares classified by the SVM classifier. Each segmentation square belonging to a specific class is assigned with the corresponding color (see Fig. 3). Basically, one class contains everything but apoptotic cells (red segmentation squares in Fig. 3) and the other everything but cancer cells (blue segmentation squares in Fig. 3. The red color segmentation squares generate an image, which discards all the apoptotic class cells (see Fig. 4b). The blue color segmentation squares does the same on the cancer class cells (see Fig. 4c). The result is the generation of two images, each of them containing one kind of cells.

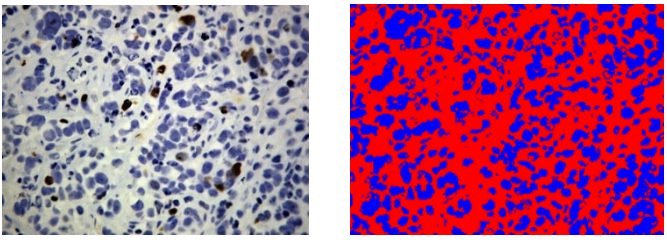

Fig. 3. Original and Segmented Image

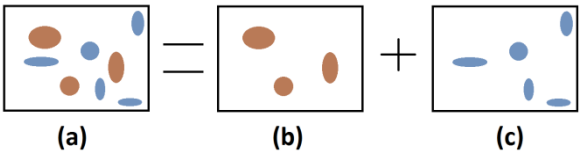

Fig. 4. a) Original Image b) Red pixels in **Fig. 3** c) Blue pixels in **Fig. 3**

2.3 Thresholding

Adaptive thresholding segmentation is applied on the 8-bit grayscale versions of each of the generated images. Thresholding is one of the most common methods for image segmentation task and it is used in several applications [14]. In order to find the optimal value for thresholding [15] the images, the following procedure is followed: An initial threshold value is set at the minimum possible value (1), which separates the foreground from the background pixels. The average values of the pixels up to the threshold value (the foreground objects) and the pixels above (the background objects) are calculated. Afterwards, threshold value is increased and the process is repeated until the threshold value is greater than the composite average. The optimal threshold for the cancer cells image, utilizing the above technique, is T1=170 from the 0-255 scale. Based on the optimal threshold value found a new image is generated by replacing every pixel whose value is lower or equal to T1=170 with black, and the pixels over that value with white. The same procedure is applied to the apoptotic image as well. Because the apoptotic cells are lighter than the dark cancer cells require higher values to the threshold in order to distinct them from the background. Therefore, the above procedure ended up with a T2=196 threshold value for the apoptotic cells image.

2.4 Noise Removal and Object Quantification

A. Majority Voting Technique

In order to enhance the segmentation results a morphological filtering is required. A simple majority vote technique described in [16] with a dynamic vote limit was utilized in order to accomplish this task. The value of the central segmentation square is set by the majority vote of its eight neighbor segmentation blocks. The limit is set to

five, which means if five or more neighbor segmentation squares have the same value, which differs from the value of the central one; it is automatically set to the same value with its neighbors.

B. Watershed Filtering

The watershed filter is then applied on this image in order to accurately cut the merged particles and provide a clear image for the quantification procedure. Majority vote technique was applied before watershed, since we need first to enhance the foreground objects and then try to separate the merged ones.

C. Size-based Correction and Labeling

Another major factor, except the color, that characterizes a cancer or apoptotic cell is its size. Based on the assistance of our expert pathologists, the size of a valid cancer cell is interpreted as an approximately 100±15 pixels area on the image. For that reason an analyze particle method, similar to a component labeling [18] algorithm was developed. This developed algorithm allows user to set the minimum size of a valid countable cell. More specifically, the developed algorithm scans the image, from right to left beginning from the top and moving to the bottom, until it finds a pixel of a foreground object, in our case a black one since we deal with a binary image. It checks if that pixel belongs to an object that has been already scanned, and if it is it ignores it, otherwise it begins the scanning of a new object by registering it as the first pixel of that object. Afterwards the algorithm will check, based on the previous move of the scanner (in the first step default move is Right), from the westernmost point and moving clockwise to find a pixel of the same color, which has not been visited yet. This loop continues until there is no neighbor point to visit. In the end the entire object will be scanned in a clockwise spiral way. If the size of the object is less than the minimum it turns it to white, and practically it deletes it. This method is as accurate as the Analyze Particles method of imageJ is, plus it requires less memory and it's more simplified than the initial connected component-labeling algorithm. The time required for the proposed tool to complete the above procedures is about 48 seconds on an Intel ® Core™ i7 CPU Q720 at 1.60GHz with 8 GB RAM installed. The whole quantification procedure is depicted in Fig. 5.

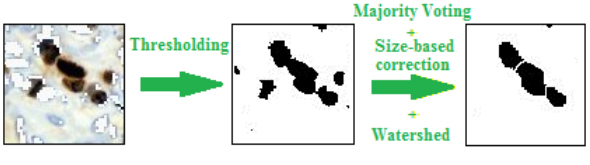

Fig. 5. Example of the proposed Quantification Procedure

3 Experimental Results

The application of the developed tool in 15 images is depicted in Fig. 6. The detection and quantification results are presented in Table 1. As depicted in Table 1, for three

images (one of each type), the proposed tool achieved an average of 96.5% accuracy. Likewise, the tool proved sufficient in the testing of all the images of the dataset achieving a 95.37% overall accuracy (see Table 2). Some miscounts may occur due to the existence of extremely merged cells (which look like one round entity), which cannot be separated by the watershed algorithm.

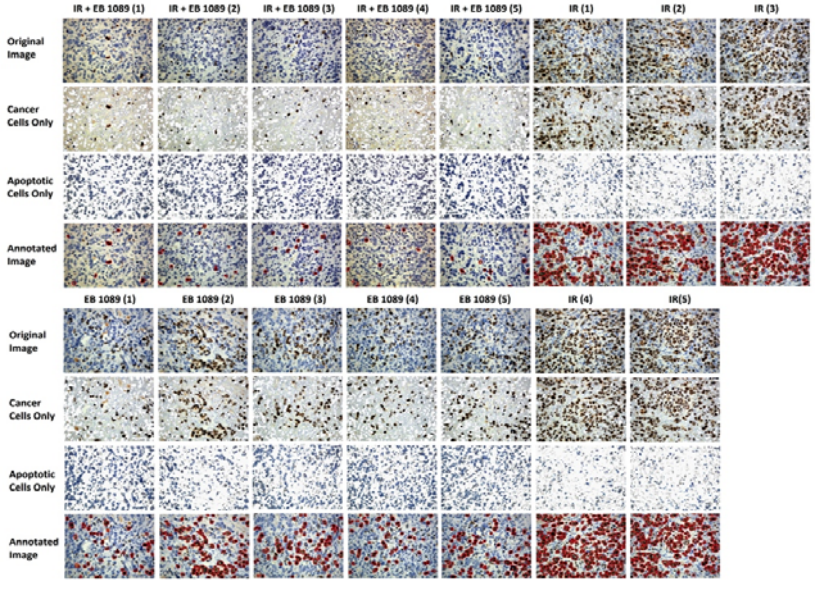

Fig. 6. Visualized Automatic Cell Recognition and Quantification Results

The performed evaluation is in coincidence with the results reported in [15], indicating also the same efficiency of each treatment. In Fig. 6, only the cancer cells are annotated in order to be easier for the physician to obtain an objective perception about the pathogenesis. An option for the annotation of the apoptotic is available as well.

Table 1. Confusion matrices of three random indicative images one of each type (IR+EB 1089(3),IR(1) and EB 1089(2))

IR + EB 1089 (3)	true Cancer cells	true Apoptotic cells	class precision (%)	Total Accuracy (%)	IR (1)	true Cancer cells	true Apoptotic cells	class precision (%)	Total Accuracy (%)	EB 1089 (2)	true Cancer cells	true Apoptotic cells	class precision (%)	Total Accuracy (%)
pred. Cancer cells	13	2	86.67		pred. Cancer cells	147	9	94.23		pred. Cancer cells	109	8	93.16	
pred. Apoptotic cells	1	263	99.62		pred. Apoptotic cells	12	212	94.64		pred. Apoptotic cells	4	203	98.07	
class Recall (%)	92.86	99.25		98.92	class Recall (%)	92.45	95.93		94.47	class Recall (%)	96.46	96.21		96.30

Table 2. Confusion matrix for the whole image dataset (Cell Recognition / Quantification)

ALL images	**true Cancer cells**	**true Apoptotic cells**	**class precision (%)**	**Total Accuracy (%)**
pred. Cancer cells	1404	132	91.41	
pred. Apoptotic cells	98	3332	97.14	
class Recall (%)	93.48	96.19		95.37

Table 3. Results of the Tool runs in treatment groups

Image	**IR + EB 1089 (1)**	**IR + EB 1089 (2)**	**IR + EB 1089 (3)**	**IR + EB 1089 (4)**	**IR + EB 1089 (5)**	**IR (1)**	**IR (2)**	**IR (3)**
Cancer Cells	7	18	13	17	11	147	168	204
Apoptotic Cells	241	265	263	278	229	212	195	165
Image	**IR (4)**	**IR (5)**	**EB 1089 (1)**	**EB 1089 (2)**	**EB 1089 (3)**	**EB 1089 (4)**	**EB 1089 (5)**	
Cancer Cells	228	197	51	109	85	58	91	
Apoptotic Cells	119	152	250	203	251	253	256	

4 Conclusions

In this paper we proposed a sufficient combination of Support Vector Machines along with Majority Voting and Watershed algorithm, in order to characterize and quantify different types of cells. The simplified connected component-labelling algorithm proved quite fast and sufficient for the quantification of the above cells. The proposed tool was applied on breast cancer biopsy images and provided a ratio of cancer cells over apoptotic cells in order to measure the efficiency of various treatments. The custom code was developed in Java; thus the tool can be easily integrated in an Application Server for web-based execution through Web services.

References

1. German, R.R., Fink, A.K., Heron, M., Johnson, C.J., Finch, J.L., Yin, D.: The Accuracy of Cancer Mortality Group: The accuracy of cancer mortality statistics based on death certificates in the United States. Cancer Epidemiology 35(2), 126–131 (2011)

2. Loncaster, J., Dodwell, D.: Adjuvant radiotherapy in breastcancer. Are there factors that allow selection of patients who do notrequire adjuvant radiotherapy following breast-conserving surgery forbreast cancer? Minerva Med. 93, 101–107 (2002)
3. Chen, A., David, B.H., Bissonnette, M., Scaglione-Sewell, B., Brasitus, T.A.: 1, 25-Dihysdroxyvitamin D3 stimulates activator Protein- 1 dependent Caco-2 cell differentiation. J. Biol. Chem. 274, 35505–35513 (1999)
4. Hansen, C.M., Hamberg, K.J., Binderup, E., Binderup, L.: Seocalcitol (EB 1089): A vitamin D analogue of anticancer potential. Background, design, synthesis, preclinical and clinical evaluation. Curr. Pharm. Design 6, 803–828 (2000)
5. Loukas, C.G., Wilson, G.D., Vojnovic, B., Linney, A.: An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. Cytometry Part A 55A(1), 30–42 (2003)
6. Saveliev, P., Pahwa, A.: A topological approach to cell counting. In: Proceedings of Workshop on Bio-Image Informatics: Biological Imaging, Computer Vision and Data Mining, Center for Bio-Image Informatics, University of California - Santa Barbara, USA, January 17-18 (2008)
7. Phukpattaranont, P., Boonyaphiphat, P.: Colour based segmentation of nuclear stained breast cancer cell images. ECTI Transactions on Electrical Eng. Electronics and Communications 5(2), 158–164 (2007)
8. Maglogiannis, I., Sarimveis, H., Kiranoudis, C., Chatzioannou, H., Oikonomou, N., Aidinis, V.: Radial Basis Function neural networks classification for the recognition of idiopathic pulmonary fibrosis in microscopic images. IEEE Transactions on Information Technology in Biomedicine 12(1), 42–54 (2008)
9. Tosun, A.B., Gunduz-Demir, C.: Graph Run-Length Matrices for Histopathological Image Segmentation. IEEE Transactions on Medical Imaging 30(3), 721–732 (2011)
10. Soule, H.D., Vazquez, J., Long, A., Albert, S., Brennan, M.: A human cell line from a pleural effusion derived from a breast carcinoma. Journal of the National Cancer Institute 51(5), 1409–1416 (1973)
11. Abramoff, M.D., Magalhaes, P.J., Ram, S.J.: Image Processing with Image. Biophotonics International 11(7), 36–42 (2004)
12. Batenburg, K.J., Sijbers, J.: Adaptive thresholding of tomograms by projection distance minimization. Pattern Recognition 42(10), 2297–2305 (2009)
13. Ridler, T.W., Calvard, S.: Picture thresholding using an iterative selection method. IEEE Trans. System, Man and Cybernetics, SMC 8, 630–632 (1978)
14. Harangi, B., Qureshi, R.J., Csutak, A., Petö, T., Hajdu, A.: Automatic detection of the optic disc using majority voting in a collection of optic disc detectors. In: Proceedings of ISBI 2010, pp. 1329–1332 (2010)
15. Sundaram, S., Sea, A., Feldman, S., Strawbridge, R., Hoopes, P., Demidenko, E., Binderup, L., Gewirtz, A.: The Combination of a Potent Vitamin D3 Analog, EB 1089, with Ionizing Radiation Reduces Tumor Growth and Induces Apoptosis of MCF-7 Breast TumorXenografts in Nude Mice1. Clinical Cancer Research 9, 2350–2356 (2003)
16. Suzuki, K., Horiba, I., Sugie, N.: Linear-time connected-component labeling based on sequential local operations. Computer Vision and Image Understanding 89(1), 1–23 (2003)
17. National Cancer Institute, `http://web.ncifcrf.gov/`