34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Cancer Cells Detection and Pathology Quantification Utilizing Image Analysis Techniques

Theodosios Goudas, *student Member, IEEE, and* Ilias Maglogiannis, *Senior Member IEEE*

***Abstract*—This paper presents an advanced image analysis tool for the accurate and fast characterization and quantification of cancer and apoptotic cells in microscopy images utilizing adaptive thresholding and a Support Vector Machines classifier. The segmentation results are also enhanced through a Majority Voting and a Watershed technique. The proposed tool was evaluated by experts on breast cancer images and the reported results were accurate and reproducible.**

I. Introduction

Several processing steps are required for the accurate characterization and analysis of biomedical image data. 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. 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 rapidly and accurately the status and the future evolution of tumor cells.

Breast cancer 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 histological imaging domain, where images obtained by optical or electronic microscopy is a scientific field where image analysis and image processing techniques may apply. There exist several related works in the literature. For instance Loukas et al [5], achieved cell counting in complex large scale histological images by utilizing edge and color information. Saveliev and Pahwa [6] approached the cell counting problem using a topological working algorithm on binary and grayscale images, suitable for different types of cells. Phukpattaranont and Boonyaphiphat [7] 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 [8] successfully classified biological microscopic images of lung tissue sections with idiopathic pulmonary fibrosis. Tosun and Gunduz-Demir [9] proposed an effective algorithm for segmenting histopathological images.

Ilias Maglogiannis and Theodosios Goudas are with the University of Central Greece, Dep. of Computer Science and Biomedical Informatics (e-mail: imaglo@ucg.gr; goudas@ucg.gr).

In this paper an advanced image analysis tool for the fast detection and quantification of the cancer and the apoptotic cells in the microscopy image is presented. The proposed tool provides physicians the feature of repeating this complex procedure for many samples in a short time. The proposed tool was used for the evaluation of the influence of the vitamin D3 analogue EB1089 [4] with fractionated radiation on growth and apoptosis human breast cancer tumour MCF-7 [10] cells injected in six week old 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.

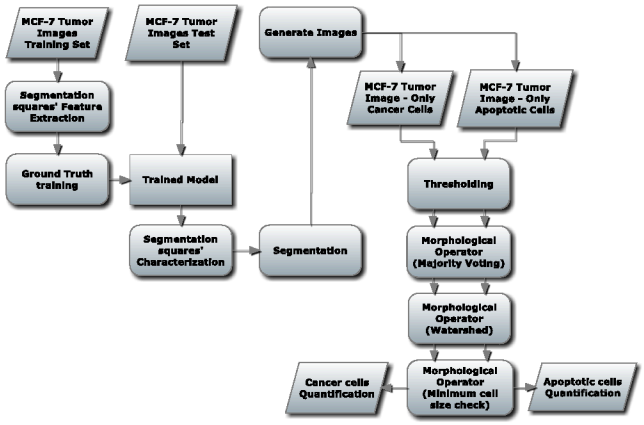

Figure 1. Data Flow Block Diagram of the Proposed Tool

II. Materials and Methods

Fig. 1 illustrates the data flow diagram of the proposed tool. After the import of an image in the system, it is first edited by the classification model, which separates the different types of cells into individual generated images. An adaptive thresholding segmentation technique is applied to enhance the visibility of the foreground objects of each generated image separately. Minor noise generated through misclassification of small segments is removed by adopting morphological operators such as Majority Voting and Watershed filtering. A size correction procedure was developed for the valid and accurate quantification of each type of cells. The training of the above model is based on the ground truth provided by expert pathologists.

A. Image Dataset Description

The microscopy images dataset was obtained from the National Cancer Institute tumor repository (Frederick, MD) [16]. These images were taken from six week old mice, which were injected the MCF-7 human breast cancer type. Subconfluent cultures, which were grown in RPMI 1640 on 37°C, 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.

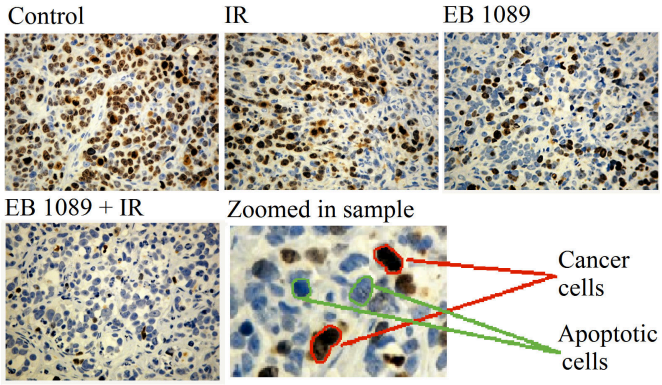

Figure 2. Four treatment groups of cancer cell images and indication of Cancer and Apoptotic Cells

When the tumor volumes reached 150–200 mm^3, 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 Fig.2, cancer and apoptotic cells have a circular form and sometimes are merged making harder their quantification.

B. Feature Extraction and Classification

Support Vector Machines (SVM) classifier was selected as the most efficient classifier, after an evaluation procedure among others, as Naïve Bayes, k-Nearest Neighbor and Decision Trees.. The mean red, green and blue channel values were utilized as image features. Each instance was labelled with the corresponding class in order to make the SVM model capable to recognise similar instances from unknown images and classify them. The ten-fold cross validation has been adopted for testing the classification accuracy. 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.

Customized code was developed for the separation of cancer and apoptotic cells from the entire image. 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 do 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.

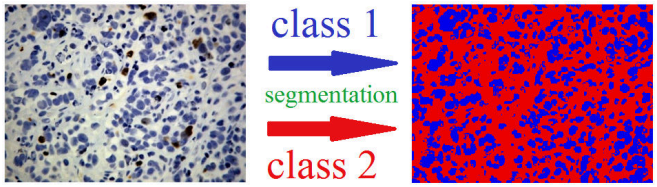

Figure 3. Original and Segmented Image

C. Thresholding

Adaptive thresholding segmentation is applied on the grayscale versions of each of the generated images [11]. In order to find the optimal value for the threshold [12], 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 on the 0-255 scale. 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.

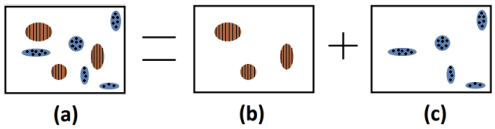

Figure 4. a) Original Image b) Red pixels in Fig. 3 c) Blue pixels in Fig. 3

D. Noise Removal and Object Quantification

1) Majority Voting Technique

In order to remove misclassification issues, a morphological filtering is required. A majority vote technique described in [13] 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 majority 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.

2) 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 firstly need to enhance the foreground objects and then try to separate the merged ones.

3) Size-based Correction and Labeling

Size is another major factor, except the color, that characterizes a cancer or apoptotic cell. Based on the assistance of our expert pathologists, the size of a true 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 algorithm [15], was developed. The developed algorithm provides the feature setting the minimum size of a valid countable cell.

Pseudocode of the component labeling algorithm

INPUT: Binary Image after Adaptive Thresholding
INITIALIZATION: set move variable to R (Right move)
FOR j = 1 TO image_height **DO**
  FOR i = 1 TO image_width **DO**
    WHILE pixel[i,j] belongs to object
      IF pixel[i,j] is the first pixel that belongs to object **THEN**
        IF pixel[i,j] has been visited before **THEN**
          Break the WHILE statement and go to the next pixel
        ENDIF
        Store i and j coordinates as the initial point of the current scanning object
      ENDIF
      i and j coordinates belong to the scanning object, store them
      REPEAT
        *Set the ud and rl coordinates starting from the westernmost neighbor and moving clockwise on each repetition until it reaches one point before the previous
        IF pixel[rl][ud] belongs to object **AND** never visited before **THEN**
          set i to rl and j to ud
          valid westernmost pixel of the object was found
          set the move to the corresponding direction of the westernmost neighbour
        ENDIF
      WHILE valid westernmost neighbor pixel of the object not found **AND** number of repetitions is less or equal to 6
      IF valid westernmost pixel of the object not found **THEN**
        End the scanning of the current object and set i, j to the initial point of the scanned object
        IF size of the scanned object is smaller than the size of a valid cell (set by the user) **THEN**
          Set as background all the pixels which belong to the current object of the scanned object
        ENDIF
      ENDIF
    ENDWHILE
  ENDFOR
ENDFOR
OUTPUT: Transformed binary image with the filtered cells and cell quantification figures for each cell type

*The assignment of the westernmost pixel, by which the clockwise search routine begins, is based on a swich-case statement, which contains all the 8 possible directions as cases for the switch statement. (see Fig. 5)

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. The algorithm checks if the specific pixel belongs to an object that has been already scanned. In case the pixel was previously scanned, it is ignored. Otherwise it begins the scanning of a new object by registering it as the first pixel of that object. Afterwards the algorithm will seek, 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. The algorithm works perfectly on 4-connected patterns, but it may have problem on scanning 8-component patterns. The average time required for the proposed tool to complete the above procedures on an Intel ® Core™ i7 CPU Q720 at 1.60GHz with 8 GB RAM installed for an image size of 512x384 pixels is 48 seconds.

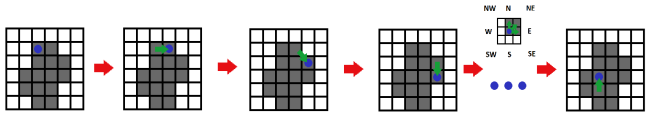

Figure 5. The labeling step of the proposed algoritmh

TABLE I. CONFUSION MATRICES OF THE THREE IMAGES (IR+EB 1089(3),IR(1) AND EB 1089(2))

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

III. 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 I. As depicted in Table I, 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 II). 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. The performed evaluation is in coincidence with the results reported in [14], 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.

TABLE II. CONFUSION MATRIX FOR ALL THE IMAGES (MASS 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

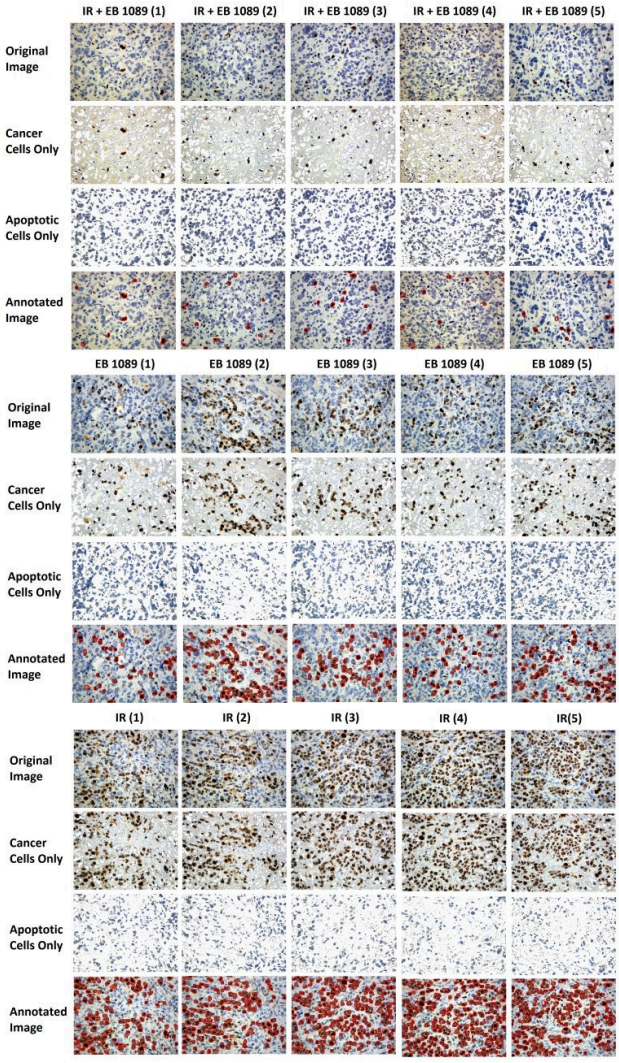

Figure 6. Visualized Automatic Cell Recognition and Quantification Results

TABLE III. RESULTS OF THE TOOL RUNS IN TREATMENT GROUPS: IR (RADIATION), EB 1089 (DRUG) AND EB 1089 + IR COMBINED

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	

IV. CONCLUSIONS

In this paper a sufficient combination of Support Vector Machines along with Majority Voting and Watershed algorithm was proposed, in order to characterize and quantify different types of cells. The simplified connected component-labeling 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] R.R. German, A.K. Fink, M. Heron, C.J. Johnson, J.L. Finch, D. Yin, "the Accuracy of Cancer Mortality Group: The accuracy of cancer mortality statistics based on death certificates in the United States." *Cancer Epidemiology*, vol. 35, Issue 2, pp. 126-131, April 2011.

[2] J. Loncaster, D. Dodwell, "Adjuvant radiotherapy in breast cancer. Are there factors that allow selection of patients who do not require adjuvant radiotherapy following breast-conserving surgery for breast cancer?" *Minerva Med*. 93, pp. 101–107, 2002.

[3] A. Chen, B.H. David, M. Bissonnette, B. Scaglione-Sewell, T.A. Brasitus, "1, 25-Dihysdroxyvitamin D3 stimulates activator Protein- 1 dependent Caco-2 cell differentiation". *J. Biol. Chem.* 274: 35505–35513, 1999.

[4] C.M. Hansen, K.J. Hamberg, E. Binderup, L Binderup, "Seocalcitol (EB 1089): A vitamin D analogue of anticancer potential. Background, design, synthesis, preclinical and clinical evaluation." *Curr. Pharm. Design*, 6: pp. 803–828, 2000.

[5] C.G. Loukas, G.D. Wilson, B. Vojnovic, A. Linney, "An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections", *Cytometry Part A*, Volume 55A, Issue 1, September, pp. 30–42, 2003.

[6] P. Saveliev, A. Pahwa, "A topological approach to cell counting." *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] P. Phukpattaranont, P. Boonyaphiphat, "Colour based segmentation of nuclear stained breast cancer cell images." *ECTI Transactions on Electrical Eng. Electronics and Communications*, vol. 5, no. 2, pp. 158-164. Aug. 2007.

[8] I. Maglogiannis, H. Sarimveis, C. Kiranoudis, A.A. Chatzioannou, N. Oikonomou, V. Aidinis, "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), pp. 42-54, 2008.

[9] A.B. Tosun, C. Gunduz-Demir, "Graph Run-Length Matrices for Histopathological Image Segmentation." *IEEE Transactions on Medical Imaging,* vol. 30, No.3 March pp. 721-732, 2011.

[10] H.D. Soule, J. Vazquez, A. Long, S. Albert, M. Brennan, "A human cell line from a pleural effusion derived from a breast carcinoma." *Journal of the National Cancer Institute* 51 (5), pp.1409–1416, 1973.

[11] K.J. Batenburg, J. Sijbers, "Adaptive thresholding of tomograms by projection distance minimization." *Pattern Recognition*. Volume 42, Issue 10, pp.2297-2305, October 2009.

[12] T.W. Ridler, S. Calvard, "Picture thresholding using an iterative selection method." *IEEE Trans. System, Man and Cybernetics*, SMC-8 pp.630-632, 1978.

[13] B. Harangi, R.J. Qureshi, A. Csutak, T. Petö, A. Hajdu, "Automatic detection of the optic disc using majority voting in a collection of optic disc detectors." *Proceedings of ISBI'2010*, pp.1329-1332, 2010.

[14] S. Sundaram, A. Sea, S. Feldman, R. Strawbridge, P. Hoopes, E. Demidenko, L. Binderup, A. Gewirtz, "The Combination of a Potent Vitamin D3 Analog, EB 1089, with Ionizing Radiation Reduces Tumor Growth and Induces Apoptosis of MCF-7 Breast Tumor Xenografts in Nude Mice1." *Clinical Cancer Research*, vol. 9, pp. 2350-2356 June 2003.

[15] K. Suzuki, I. Horiba, N. Sugie, "Linear-time connected-component labeling based on sequential local operations." *Computer Vision and Image Understanding*, vol. 89 Issue 1, pp.1-23, January 2003.

[16] National Cancer Institute, http://web.ncifcrf.gov/