COMPUTER-AIDED EVALUATION OF HER-2 STATUS IN FLUORESCENT IN SITU HYBRIDIZATION IMAGES

F. Raimondo, M. A. Gavrielides, G. Karayannopoulou, K. Lyroudia , I. Pitas, I. Kostopoulos

Aristotle University of Thessaloniki, Greece

ABSTRACT -- The analysis of Her-2/neu status is an effective indicator for the diagnosis of several types of breast carcinomas. Conventional evaluation is a difficult task since it involves manual counting of dots in multiple fluorescent in situ hybridization (FISH) images. In this paper we present a multistage algorithm for the automated evaluation of Her-2/neu status by the analysis of FISH images from breast carcinomas. The algorithm focuses on the detection of FISH spots and on the cell nuclei segmentation in order to perform overall case classification as positive or negative. Spots detection includes mainly a top-hat filtering stage, a binary thresholding, a 3D template matching and a grey level contrast evaluation. Nuclei segmentation consists of a non-linear blue channel correction step, a global thresholding by Otsu algorithm, a grey level hole classification by a geometric rule and of the marked watershed transform using local h-dome maxima as markers. By the measurement of the FISH signals ratio per cell nucleus we perform the classification of cases. The performances of the algorithm were evaluated with receiver operating characteristic (ROC) analysis.

Keywords: fish, breast carcinomas, spot detection, nuclei segmentation, case classification.

I. INTRODUCTION

The HER-2/neu (c-erbB2) oncogene is a tyrosine kinase receptor that is overexpressed in approximately 20-30% of high-grade invasive breast carcinomas and has been shown to be a valuable prognostic indicator [1]. HER-2 positive tumours can be more aggressive and their status can predict response to targeting therapy with trastuzumab (Herceptin) monoclonal antibodies and adjuvant chemotherapy. Knowing that a cancer is HER-2/neu positive helps a medical team select the appropriate treatment. Overexpression of the protein product of HER-2/neu gene is usually a consequence of gene amplification, in which multiple copies of the gene appear through the genome. It is thus possible to determine HER-2/neu status by analyzing the numbers of gene copies centrally or the amount of protein peripherically. Currently, the two most widely used technologies to determine HER-2/neu status are immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH). IHC uses specific antibodies to stain proteins (products) *in situ,* which allows the identification of many cells types that could be visualized by classical microscope. FISH imaging allows selective staining of various DNA sequences and thereby the detection, analysis and quantification of specific numerical and structural abnormalities within these nuclei. The IHC test measures the protein coded by the HER2 gene, whereas FISH measures the number of copies of the HER-2 /neu gene present in the tumor cells. A recent study by Bartlett et. al. [2] considered the accuracy, reproducibility and availability of different techniques for the evaluation of Her-2/neu status and recommended screening by immunohistochemistry followed by FISH testing of cases with intermediate staining intensity (cases scored 2+according to Hercep test). They suggested that the use of automated analysis may increase testing precision and predicted a wider future use of FISH analysis as a more cost-effective technique.

The process of evaluating HER-2/neu status from FISH images involves the manual scoring of the ratio of HER-2/neu over CEP 17 dots within each cell nucleus and then averaging the scores for a number of cells in the order of 60. Several images usually need to be read to reach the desired number of dot-including nucleus. A ratio of >= 2.0 of HER-2/neu to CEP 17 copy number denotes amplification. The reading of FISH images is a difficult task since manual dot scoring over a large number of nuclei is a time consuming and fatiguing technique. The automation of the process has been suggested as a way to increase cost-effectiveness and reduce inter-observer variability in reading FISH images.

Several methods have been proposed for the automated evaluation of FISH signals, even though they were not applied directly for measuring Her-2/neu gene amplification of breast samples. Most methods focused on automatic spot counting whereas only very few focused on case-based classification of FISH images. Netten et al. [3] focused on automatic counting of dots per cell nucleus in slides of lymphocytes from cultured

blood. Their method consisted in selecting regions of interest containing at least one nucleus and in a spot detection using the top hat transform and a nonlinear Laplacian filter. Solorzano et al. [4] developed a method to study leukocytes in blood samples. They segmented nuclei using the ISODATA thresholding algorithm. Then, the watershed algorithm incorporating the distance transform was used to isolate nuclei and FISH dots were detected using the top hat transform. Kozubek et al. [5] developed a system that acquired 2-D and 3-D FISH images. For 2-D analysis, the system segmented the nuclei using bimodal histogram thresholding and morphological features for further binary image processing. Then, the system detected hybridized dots within each segmented nucleus using a watershed-based algorithm. 3-D analysis was performed by analyzing the pre-extracted nuclei and dot features for sequential 2-D slices. Lerner et al. [6-9] proposed a FISH image classification system based on the properties of in- and out-of-focus images captured at different focal planes. In a later study, the authors employed a Bayesian classifier instead of an NN, to avoid dependency on a large number of parameters.

In this paper we present a multistage algorithm for the automated evaluation of HER-2/neu status in breast carcinomas samples. It includes two parallel stages for nuclei segmentation and dot detection respectively. It takes into account multiple images of a specific case and quantifies the HER-2/neu status in a collective manner. Moreover, the system is evaluated with ROC analysis. The paper is organized as follows: Section II describes the database used for the development and evaluation of the method. The method is detailed in section III whereas Section IV presents the evaluation results and a related discussion. Finally, conclusions are stated in Section V.

II. MATERIALS:

Paraffin sections of 4μm thickness were incubated overnight at 60°C. Deparaffinization, pretreatment, enzyme digestion and fixation of slides were performed using the Vysis Paraffin Pretreatment kit according to the manufacturer's recommended protocol. After proteolysis, tissue sections were denaturated at 85°C for 2 minutes, then the PathVysion HER-2 DNA Probe (LSI HER-2/CEP17 probe, Abbott GmbH and Company, KG, Wiesbaden-Delkenheim, Germany) was added and hybridization took place at 37˚ C in a moist chamber for 14-18h (overnight incubation). The following day the slides were washed with post-hybridization buffer (2X SSC and 0,3% NP-40) at 72°C for 2 minutes, followed by counterstaining of the nuclei with 4, 6-diamino-2phenylindole dihydrochloride (DAPI). For each specimen, at least 60 non overlapping nuclei were scored for both Her-2/neu and chromosome 17 signals by image analysis. Hybridization signals were enumerated utilizing a Zeiss, Axioskop 2 HBO 100. Her-2/neu gene amplification was determined by a ratio of Her-2/neu gene copies to chromosome 17 centromeres. According to the manufacturer's recommendations the cases with a ratio ≥2 were determined as amplified, while those having a ratio <2 as not amplified.

III. METHOD:

The algorithm for the classification of FISH images was based on the accurate measurement of red/green spot ratio (corresponding to the ratio of HER2/CEP 17) per cell nucleus. For that reason, two parallel stages for spot detection and cell nuclei segmentation were developed as described below.

A. FISH SPOT DETECTION:

Despite the fact that the main content of FISH image red and green channels is constituted of spots, many FISH images frequently contain noisy areas consisting of large stains. For this reason, a preprocessing step for noise removal is needed. An effective solution is provided with top-hat filtering, as was proposed in [10]. A disk of 4-pixel radius was chosen as the structuring element of the top-hat transform.

A typical grey level histogram of the top-hat output presents a strong unimodal trend consisting of a peak followed by a very steep monotonous decrease and a second flat part. A modification of the algorithm proposed in [11] was used to estimate two thresholds for the top-hat red and green channel output respectively. The algorithm assumes that there is one dominant mode in the image histogram. A straight line is drawn from the peak to the high intensity end of the histogram. More precisely, the line starts at the largest frequency bin A and finishes at the first empty bin B of the histogram following the last filled bin. The threshold is selected as the histogram index *Th* that maximises the perpendicular distance between the line and histogram curve. This procedure is illustrated in Fig.1.

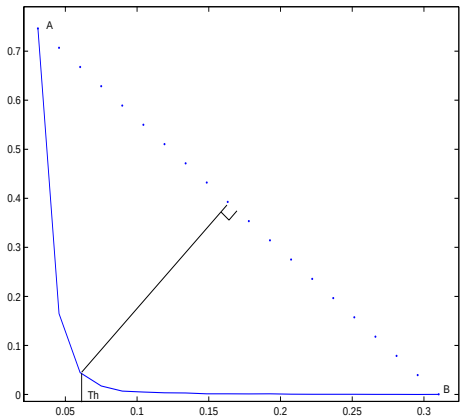


Fig.1: Typical grey level histogram of top-hat output and threshold selection.

For this application, we modified this algorithm by applying it not on the entire image histogram, but rather on the pixels belonging to the last k bins of the histogram. The value of k is estimated in the following way: if we indicate with N the total number of pixels of the image and with $h(i_n)$ the intensity image histogram value relative to the $n-th$ bin, then we consider the last k bins such that:

$$\sum_{n=N-k}^{N} h(i_n) \geq N \cdot p \qquad (1)$$

The value of p was chosen equal to 2.5%. The histogram resulting from the selected pixels still presented a strong unimodal trend. The resulting threshold is proven to be rather insensitive to the value of p .

Even the best threshold choice is not enough to isolate all true spots from false ones using only the red or green channel intensity. For this reason, we measure the similarity between every candidate spot and two spot templates, one for the red channel and the other for the green channel. To measure the similarity we use normalized cross correlation [12]. The estimation of the spot shape template is performed on training FISH images in the following way: considering the red and green channel independently, we estimate the center of every red and green manually labeled spot as the pixel with maximum channel intensity. A 7×7 window positioned on every spot center is saved as a template for spots in the red and green channel respectively. Two spot template windows T_G and T_R are estimated by averaging the respective spot channel intensities:

$$T_R(x,y) = \frac{1}{N_R} \sum_{i=1}^{N_R} f_{R_i}(x,y) \qquad (2)$$

$$T_G(x,y) = \frac{1}{N_G} \sum_{i=1}^{N_G} f_{G_i}(x,y) \qquad (3)$$

where N_R and N_G are the number of used red and green spots, $x = 1,...,7$ and $y = 1,...,7$ are coordinates in a 7×7 window and f_{R_i} and f_{G_i} are red and green channel intensities of $i-th$ spot image.

The obtained normalized cross correlation C_R and C_G were evaluated only for FISH image positions where a) the top-hat output is above threshold and b) a nuclei is present, as found by the nuclei segmentation procedure to be described subsequently. In order to select red/green spot positions, two positive thresholds Th_R and Th_G are used; spots with a value of C_R and C_G lower then Th_R and Th_G respectively, are discarded while the remaining ones are used as input for the next selection step of the FISH spot detection algorithm.

Finally**,** for every detected spot from the previous step, a channel intensity contrast measure is used. This further processing step is performed to discard spots whose shape is very similar to the template one, but have a low channel intensity contrast with respect to their surrounding pixels, making them appear invisible to the human eye. The contrast measure is performed using the information of the red, green and blue channel. For each spot, two vectors $\boldsymbol{v}_{for}$ and $\boldsymbol{v}_{back}$ are created. Each of the three components of vector $\boldsymbol{v}_{for}$ is estimated considering the average channel intensity of the pixels of a 5×5 window positioned on every spot center, while each of the three components of vector $\boldsymbol{v}_{back}$ is estimated considering the average channel intensity of the pixels surrounding the previous window. Then the contrast measure C_M is calculated as:

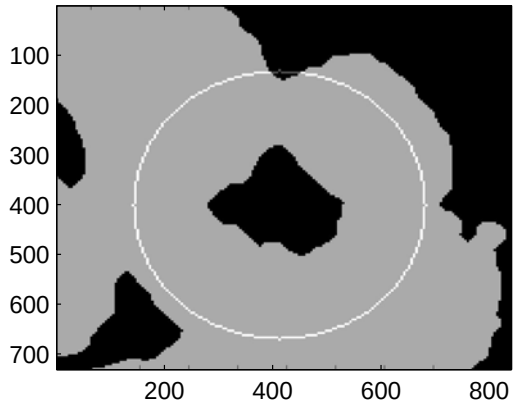
$$C_M = \frac{\left\| \boldsymbol{v}_{for} - v_{back} \right\|}{\left\| v_{back} \right\|} > T_{CM} \qquad (4)$$

The thresholds Th_R , Th_G and T_{CM} are empirically chosen in order to minimize spot classification error over the FISH images used for training.

B. CELL NUCLEI SEGMENTATION:

Cell nuclei segmentation is performed on the FISH image blue channel. For many images, cell nuclei contain inhomogeneous blue channel intensity. To overcome this problem, a nonlinearity correction step incorporating the square root function is applied in order to reduce the gray level difference between dark regions and more illuminated ones. Moreover, gray level peaks due to the presence of spots are made less intense by applying the opening morphological operator to the FISH image blue channel using a disk of 4-pixel radius as structuring element. A further top-hat filtering using as structuring element an 80-pixel radius disk is also performed to reduce the blue channel intensity of big stains.

Based on the fact that the histogram of the resulting image has a characteristic bimodal shape, we employ the algorithm by Otsu et al [13] to determine the threshold for initial nuclei segmentation. The binary image resulting from thresholding sometimes contains holes even in a single nucleus body region. This kind of holes has to be filled to enable correct nuclei segmentation. On the contrary, holes present in inter nuclei zones of overlapping nuclei should not be filled. The two types of holes are illustrated in Figs 2 a and b respectively. A method to distinguish between these two types of holes is described below.



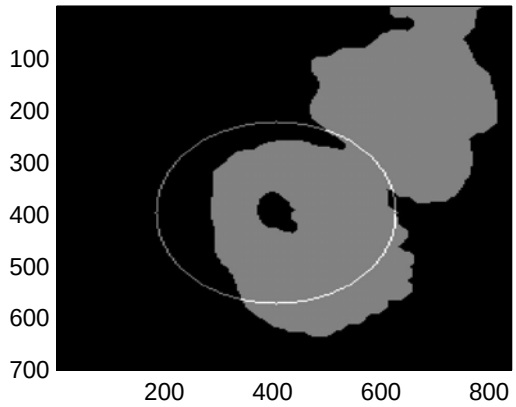
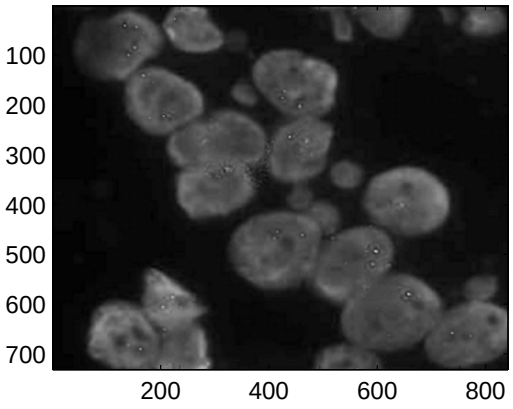


Fig.2: Top: hole due to overlapping nuclei (inter nuclei case); Bottom: nucleus body hole.



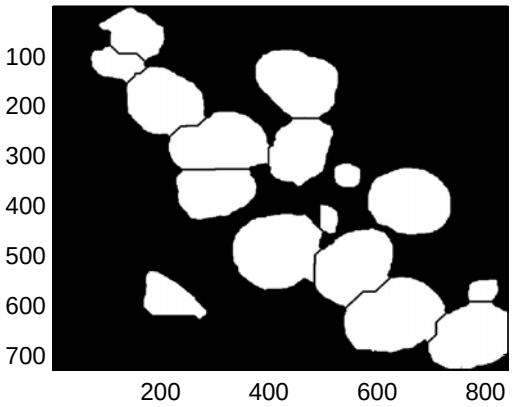


Fig.3: Top: original blue channel image; Bottom: final output of the segmentation algorithm.

As the two types of holes can not be separated into two classes using just gray level or morphological features, a geometric approach is employed. Let P be the percentage of the perimeter pixels of a circle of radius R centered around a hole centroid that are contained in the nucleus region. It can be observed from Fig. 2, that the value of P is much higher for the first type of hole (inter nuclei) than the second type of hole (nucleus region). The radius R used is slightly bigger than the average nuclei radius. It was found experimentally that the value for P varied in the range of 90% to 40% for holes of the first type and second type respectively.

The last step of the nuclei segmentation algorithm involves the marked watershed transform [14] which is employed to detect borders in overlapping nuclei clusters. The distance transform [15] is first applied to the binary image obtained from the previous step. For every pixel, this transform produces an intensity value proportional to its distance from the closest background pixel. The uneven shapes of binary object borders cause spurious local maxima of the distance transform output that do not correspond to nuclei centers. If all these maxima were used as markers, an over-segmentation would be obtained. In order to reduce this effect we calculate *h-domes* of resulting image [16]. We can define the *h-dome* as a region of pixels wherein every pixel has an intensity value greater than any of the pixels surrounding the region and the maximum intensity difference between two pixels in the region is smaller than or equal to h. Characteristic values of h are in the range $[0.5 : 2]$. This feature is employed in order to discard insignificant local maxima present in the distance transformed image. Fig.3 shows examples of the original blue channel image and the final output of the nuclei segmentation step. We note that the nuclei touching the image border are removed because they are not considered in the spots per cell counting.

The results of the spot detection and nuclei segmentation steps were combined and the ratio of red/green spots per nucleus was calculated as will be described in the next section.

IV. RESULTS

The algorithm was evaluated with respect to three different tasks: spot detection, cell nuclei segmentation and case-based classification as will be described. Receiver Operating Characteristic (ROC) curves were used to describe the performance of each task, as will be shown below.

A. SPOT DETECTION:

In order to estimate the performance of the algorithm for detecting FISH spots, a testing set of 40 FISH images was used. The true location of 887 red spots and 751 green spots was labeled by an expert. The same expert identified the location of 385 true red spots and 334 true green spots in 18 different FISH images that were employed as training set to estimate spot shape templates. The false positive rate was defined as the ratio between the total number of detected spots not present in the ground truth over the total number of detected spots. The true positive rate was defined as the ratio between the total number of correctly detected spots over the total number of spots present in the ground truth. ROC curves were constructed by collecting pairs of sensitivity (or true positive rate) and false positive rate for different thresholds. Twenty one points of the curve were estimated by varying the threshold applied to top-hat output between 0 and 1 with a step of 0.05 while fixing the thresholds of normalized cross correlation, Th_R and Th_G, and threshold T_{CM} of gradient intensity value to 0.7, 0.65 and 0.3 respectively. Curves were estimated for red and green spots separately and are displayed in Fig. 4.

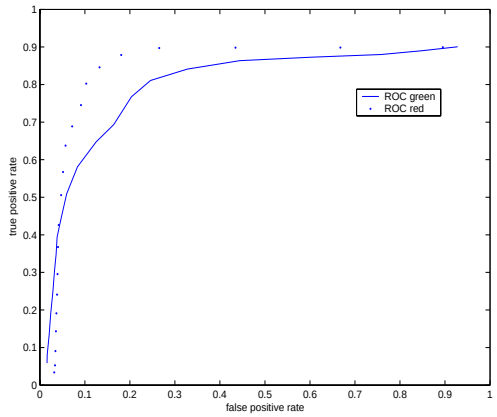


Fig.4: ROC curves relative to red and green spots detection.

It can be seen from Fig.4 that the performance of the algorithm is noticeably better for the detection of red spots. This can be due to the fact that red spots have a more well-defined shape compared to the green ones. As a demonstration point, the algorithm can reach a sensitivity of about 92% and 80% for red and green spots respectively at a false positive rate of about 25%.

B. CELL NUCLEI SEGMENTATION:

In order to evaluate the performance of the algorithm for cell nuclei segmentation, the ratio between the area of intersection of segmented nucleus with true nucleus region over the area of the union of two regions was calculated. The ground truth for the correct nucleus boundaries was determined manually so that 1439 nuclei were labeled. ROC curves were constructed by varying a threshold for the ratio of intersection over union as described above. Fig.5 shows the resulting ROC curves for three values of h {0.5, 1.0, 1.5}.

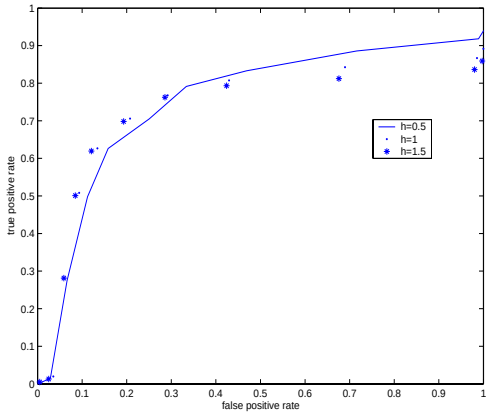


Fig.5: ROC curves relative to the cell nuclei segmentation.

As we can see performances are not much sensitive to the value of *h*, even if a value of h = 1 shows better results in the high sensitivity part of the ROC curve. It has to be noted here that for this application it is more crucial to not discard a true nucleus than to avoid merging overlapping nuclei since dot counting per nucleus is averaged and it is not overly sensitive to two nuclei being counted as one. On the other hand, correct nuclei segmentation definitely reduces the effect of background noise pixels that can be mistaken for FISH dots and effect the dot ratio calculations.

C. FISH CASE-BASED CLASSIFICATION:

Twelve patient cases, six of which were previously classified by an expert as positive and six that were classified as negative, were employed to evaluate the precision of the algorithm in classifying the different cases. For every case a certain number of images were

available. For every segmented nucleus where at least one red spot was present the ratio d was calculated, defined as:

$$d = \frac{N_R}{N_G} \qquad (5)$$

where N_R and N_G are, respectively, the number of red and green spots present in the segmented nucleus. In the case that the number of green spots is zero, we set N_G equal to one. For each case the histogram of d was calculated in order to estimate the probability $P(d \geq 2)$ of having a ratio red/green greater or equal then two. The values of $P(d \geq 2)$ for each of the testing cases were given in Tab. 1.

Table 1. Values for the estimation of the probability to have a ratio greater or equal to 2 for each of the testing cases.

Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6
0.531	0.663	0.525	0.488	0.515	0.538
Neg. 1	Neg. 2	Neg. 3	Neg. 4	Neg. 5	Neg. 6
0.239	0.319	0.287	0.201	0.323	0.245

It can be seen from Table 1 that all cases can be correctly classified as either positive or negative (100% sensitivity with 0 false positives) by using a proper threshold for the statistic $P(d \geq 2)$. These preliminary results show that a fully automated method can accurately distinguish between normal and abnormal breast tissue samples. A larger database of FISH images of breast tissue is being prepared in order to examine how well these results can generalize in a broader population.

V. CONCLUSIONS

We have developed a method for the automated evaluation of Her-2/status in breast samples by FISH image analysis. The method uses two parallel multistage algorithms, the first one for the detection of the red and the green spots and the second one for the cell nuclei segmentation. The outputs of the two algorithms were merged for estimating the average red/green ratio per cell nucleus. The performance of the proposed method was evaluated using ROC curves both for the detection of the red and the green spots and for the cell nuclei segmentation. Moreover, the overall algorithm performance for case-based classification of FISH images showed the ability of the system to distinguish between positive and negative cases. The evaluation results were encouraging for the further development and evaluation of this method.

ACKNOWLEDGEMENT

This work was supported by the EU project Biopattern: Computational Intelligence for biopattern analysis in Support of eHealthcare, Network of Excellence Project No. 508803.

REFERENCES

1. P. P. Osin and S. R. Lakhani, “The pathology of familial breast cancer. Immunohistochemistry and Molecular analysis”, *Breast Cancer Research*, vol. 1, no.1, pp. 36-40, 1999.
2. J. Bartlett, E. Mallon, and T. Cooke, “The clinical evaluation of Her-2 status: which test to use?”, *Journal of Pathology*, vol. 199, pp. 411-417, 2003.
3. H. Netten, I. T. Young, L. J. van Vliet, H. J. Tanke, H. Vroljik, and W. C.R. Sloos, "FISH and Chips: Automation of fluorescent dot counting in interphase cell nuclei". Cytometry, col. 28, pp. 1-10, 1997.
4. C. O. de Solorzano, A. Santos, I. Vallcorba, J-M Garcia-Sagredo, and F. del Pozo, "Automated FISH spot counting in interphase nuclei: Statistical validation and data correction", *Cytometry*, vol. 31, pp. 93-99, 1998.
5. M. Kozubek, S. Kozubek, E. Lukasova, A. Mareckova, et. al., "High-resolution cytometry of FISH dots in interphase nucleus nuclei", Cytometry, vol. 36, pp. 279-293, 1999.
6. B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hulten, C. M. Bishop, "Feature representation and signal classification in fluorescence in-situ hybridization image analysis", *IEEE Transactions on Systems, Man, and Cybernetics*, Part A 31(6), 655-665, 2001.
7. B. Lerner, "Bayesian fluorescence in situ hybridisation signal classification", *Artificial Intelligence in Medicine*, 30(3): 301-316, 2004.
8. B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hulten, and Christopher M. Bishop,
"Automatic signal classification in fluorescence in-situ hybridization images, *Bioimaging,* vol. 43, pp. 87-93, 2001.
9. W. F. Clocksin and B. Lerner, "Automatic Analysis of Fluorescence In-Situ Hybridisation Images", BMVC 2000.
10. F. Meyer, “Iterative Image Transformations for an Automatic Screening of Cervical Cancer”, *Journal of Histochemistry and Cytochemistry,* 27, 1979, 128-135.
11. P.L. Rosin, "Unimodal thresholding", *Pattern Recognition*, vol. 34, no. 11, pp. 2083-2096, 2001.
12. R.C. Gonzales, R.E. Woods, *Digital Image Processing*, Addison Wesley, 2nd edition, 2003.
13. N. Otsu, “A thresholding selection method from graylevel histogram”, *IEEE Trans. on Systems, Man and Cybernetics*, 9 (1):62–66, 1979.
14. S. Beucher, and F. Meyer, “The morphological approach to segmentation: the watershed transformation”, in E.R. Doughterty (Ed.),

Mathematical morphology in image processing, (New-York: Dekker, 1992).
15. Pratt, William K., *Digital Image Processing*, New York, John Wiley & Sons, Inc., 1991.
16. L.Vincent,"Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms", *IEEE Trans. Image Processing*. 2**,** 176-201, 1993.