ORIGINAL RESEARCH PAPER

Computer Science

KEY WORDS: Cervical cancer, cell segmentation, nucleus, cytoplasm

CERVICAL CELL SEGMENTATION TOOL

Abrar Ali Sheikh Department of Computer Science & IT Bhaderwah Campus University of Jammu

ABSTRACT

Cervical cancer is one of the most common malignancies among women. According to world health organization an estimated one million women worldwide are living with cervical cancer. With no symptoms at all in its pre-cancerous stage it is hard to detect and when the symptoms are visible, it is unresponsive to treatment. Fortunately, it is possible to detect it in pre-cancerous stage using pap smear test, so it could be treated. Human based smear analysis is tedious, time-consuming and error prone. Therefore, machine assisted automated screening brings significant benefits in the field. The screening of cervical cell requires detection of any morphological and structural change. An effective segmentation algorithm is needed to detect the contours of cytoplasm and nucleus and then response is feed to an automated system to diagnosis any unpropitious condition. The author in this paper builds an automated cervical cell segmentation tool in **MATLAB R2015a**. The tool intended to segment the available cervical cell image into two sub image. Out of these two sub image one will contain the extracted cytoplasm and other will contain the nucleus. The tool build will allow the user to select the image and then it will automatically segment it.

INTRODUCTION

Cervical cancer is a malignant tumor that occurs when cervical tissue cells begin to grow and replicate abnormally without controlled cell division and cell death. In such a state, the body is unable to use and manage such cells for carrying out their usual function resulting these cells transforming into a tumor. If the tumor is malignant, its cell flow through the blood stream and spread to other parts of body, as a result those parts also become infected. Usually the cervical cancer takes number of years to develop. These infected cells are then distinguished as cervical intra-epithelial neoplasia (CIN) or cervical dysplasia. The cells over the surface of cervix that show unusual changes & potentially precancerous developments are called CIN. In most of the cases CIN remains stable, or these are eliminated by host's immune system response. Although, a small percentage of cases progress to become cervical cancer, if not treated. Studies have found that CIN usually results from a virus called human papillomavirus (HPV) which is generally sexually transmitted. Although there are more than 120 types of HPV known [1], only 15 are classified as high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82) [2], 3 as probable-high-risk (26, 53, and 66), and 12 as low-risk (6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, and CP6108). In many cases even after getting infected with HPV, it is generally eliminated by the response of the host's immune system, but in many cases, where HPV is not done away with by the immune system, it may develop into cervical cancer. The common risk factors linked with cervical cancer include first intercourse at an early age, pregnancy at early age, having sex with multiple partners, weak immune system, smoking, use of oral contraceptives, improper menstrual hygiene etc. At an early stage the cervical cancer may be completely asymptomatic. The early stages of the cervical cancer are usually asymptomatic but symptoms do appear with the progression of pre-cancer to invasive cancer and typically shows abnormal vaginal bleeding, vaginal discharge, pain during vaginal intercourse. New bleeding may be experienced by the women who have had their menopause. Cervical cancer is the second most commonly diagnosed [3] and fifth deadliest [4] cancer in women throughout the world. In developing countries cervical cancer has a major share in cancer mortality [5]. Every year about 500,000 new cases are diagnosed and among which about 250,000 patient die. Because of poor access to screening and treatment services, approximately 80% of this disease occurs in women living in low- and middle-income countries [6].

Screening of cervical cancer

Screening of cervical cancer test the presence of malignant cells in cervical tissues. Detection of Cervical cancer in its initial stage assists it to be treated and avoided easily. If left untreated it develops to an invasive phase. Three different types of screening tests are currently available: I) conventional

Papanicolaoutest and Liquid Based Cytology (LCB) ii) HPV testing for high-risks HPV types iii) Visual Inspection with Acetic acid (VIA). The first and second are cellular level screening and third one is tissue level screening. [7] HPV-DNA and visual inspection tests are not possible to interface in real time because of its genetic material analysis and subjectiveness.The Papanicolaou test (Pap smear) has been the widely used method in cervical cancer screening for many decades and has showed a dramatic lowering of incidents of cervical cancer and hence in related mortality rates in many countries [5].In taking a Pap smear, cells are scraped from the outer opening of the cervix for microscopic examination and to lookup for irregularities. The aim of the test is to detect any pre-cancerous or potentially precancerous alterations called cervical intraepithelial neoplasia (CIN) or cervical dysplasia. Pap test is also used to detect endocervix and endometrium abnormalities and infections. In many developed countries, regular Pap smear screening is highly recommended for females who have had frequent sex with multiple partners. If any unusual findings are observed the test may need to be repeated within a year. If the abnormality observed requires closer examination, a detailed cervical inspection by colposcopy may be done. HPV DNA testing may also be suggested to such patients, which acts as a supplementary to Pap smear testing. Once the sample is obtained, Papanicolaou technique is used to stain it. Staining using this technique helps to differentiate the cells in smear preparation from various other bodily secretions as unstained cells cannot be seen under a simple compound microscope. Most of the abnormal results are mildly abnormal (called low-grade squamous intraepithelial lesion (LSIL)) which indicates HPV infection. Most low-grade cervical dysplasia relapse on their own without usually causing cervical cancer, but presence of dysplasia can act as a warning that greater monitoring is needed. Generally, some Pap results are high-grade squamous intraepithelial lesion (HSIL), and very few of them indicate cancer.

The review presented in this work is focused on to conventional pap smear test which is commonly used to screen out abnormal cervical cells. The cervical cells which are in pre-cancerous period are called dysplastic. It undergoes three phrases: mildly dysplastic, in which nucleus becomes larger and brighter than normal one, moderate dysplastic, in which nucleus is larger and darker, severe dysplastic, in which both nucleus and cytoplasm change size

and texture. [8]

Smear analysis is an error prone, time consuming and tedious job and lack of pathologist make it worse. Automating the process of pap smear in diagnosing bring significant benefits. The automated system would require to replace the manual activities of pap smear process into an expert system of diagnosing the unassociated data from the normal one. The manual screening begins with stacking of screening data which is then subjected to analysis by human expert and based on his knowledge the medical experts reach on the conclusion. The process of manual screening is highly dependent on expert's skills and can be limited by his subjectiveness and variability. The machine replication of pap smear analysis, requires a computational system capable of simulating the behaviour and knowledge of an expert. Such intelligent system replaces the analysis and decision making of an expert with feature extraction and classification. The approach of computer aided screening could be divided into four segments: Data enhancement, Feature extraction, Feature Selection and Classification. [9]

Data Enhancement/Pre-processing: The cervical screening data can be obtained in two form spectrum and image. The accessible data is then subjected for improvement using techniques of noise reduction and segmentation. The techniques like Savitzky-Gola (SG) is commonly used for noise diminution in spectroscopy, the limitation of SG filter is bridle by filtering techniques like Binomial and Chebyshev. For image, noise is the undesirable variability of intensity, colour and brightness which meld with image while seize of data [9]. This extraneous data will impact overall analysis and cause false diagnosis. Various noise reduction techniques have been explored and presented in this review for finer inspection.

Segmentation: Once inceptive data is revamp segmentation of region of interests is proceeded. The process of segmentation is key step as it can enhance the efficiency of screening process considerably. Number of segmentation algorithm are published and applied for screening of cervical images, and can be broadly categorised into three classes: i) Characteristic feature thresholding or clustering ii) edge detection and iii) region extraction. [10]

Feature extraction: The digital images can be explored to have up to five dimensions: three spatial, one spectral and one temporal dimension. In this review we will limit our discussion to spatial and spectral dimension. The thorough analysis of these dimensions supplies significant information about the imaged object. Some feature express only the spatial arrangement, called morphology. The feature that expresses only the optical value called densitometry. And the final complex feature of combination of spatial distribution of grey values (intensity value per pixel) called textural and structural feature [11]. Designing of feature extraction model focus on those features that can individuate the intent cells form the sample. The interested features appear to rely on the visual experience for determining the region of interest. The previous studies in the context can also provide an insight relevant to extraction of significant feature. Plenty of publications have suggested disparate computational models with quantitative methods like Seeded Region Growing Feature Extraction (SRGFE) etc., for extraction of relevant features.

Feature selection: Feature extraction enumerates a number features correlated to problem concerned. In the preceding steps of feature extraction, feature selection is intended to group the most relevant features and remove the noisy one, for input to classification system. The reduction of number of inputs to classification system ameliorates the process of training and rise the accuracy. [12] The significance of feature selection led to publication of various algorithms to address the issue. Out of many available techniques Sequential Forward Floating Selection (SFFS) by Pudil et al. (1994) shows influence over all others. [13]

Classification: Classification involves discerning cells based on features selected in the previous phase. An intelligent classification system with selected features as input reduce the burden of pathologist (if done manually) considerably and improve the result of diagnosis. The research done so far dispense has plenty of classifiers for screening of cervical cells. The accepted ones include Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, K- Nearest neighbour, Linear Discriminant Analysis (LDA), Decision Tree and many others. [9]

Out of all these phases, Segmentation is pivotal one as the morphological features of nucleus determine the cell to be normal or atypical one. Hereafter the paper is intended to enunciate the literature review of research done in the paradigm of segmentation in screening of cervical cancer.

Segmentation of cervical cells

The robust, unsupervised segmentation of Papanicolaou (Pap) stained cervical cell nuclei images has been a long-standing problem in image analysis that has drawn the attention (and frustration) of research groups over the world and over the years. Perhaps the main reason for this, and the inadequate success rates that have been reported in the past, is due to the fact that it is a deceptively difficult problem. Often the images appear trivial to segment and indeed the most basic global thresholding technique will produce good results in some examples, but the Pap smear screening application requires high degrees of accuracy over extremely large data sets and this is a much more difficult problem. This situation has undermined many attempts to produce an accurate automated cervical cancer screening system, despite the range of encouraging results in the feature extraction and classification stages for this application and has led to the realization that the segmentation stage is the key to a working machine. [14]

Methodology

The fabricated tool presented in this article divide the problem of cell segmentation into two parts

i) Cytoplasm Extraction ii) Nucleus Extraction.

i) Cytoplasm Extraction Algorithm: This part of algorithm is further sub divided into six stages which include: a) Noise Removal b) Contrast limited adaptive histogram equalization (CLAHE) c) Global Thresholding d) Watershed segmentation e) Finding the largest connected component f) Final image. Figure I illustrate the detailed procedure for cytoplasm extraction algorithm with a specific image from the considered database.

a) **Noise Removal:** Read the image and apply median filtering for noise removal. Removing of noise will helps us to smoothen the image. For noise removal 2d median filter is applied. Median filtering performs median filtering of the image A in two dimensions. Each output pixel contains the median value in a 3-by-3 neighborhood around the corresponding pixel in the input image. medfilt2 method(matlab) pads the image with 0s on the edges, so the median values for points within one-half the width of the neighborhood ([m n]/2) of the edges might appear distorted.
b) **Contrast limited adaptive histogram equalization (CLAHE):** Noise less original image is enhanced using Contrast limited adaptive histogram equalization. Adapthisteq(I) method (MATLAB) enhances the contrast of the grayscale image I by transforming the values using contrast-limited adaptive histogram equalization (CLAHE)
c) **Global Thresholding:** Apply global thresholding using Otsu's method on each color channel which convert

image into binary grid of pixels. Graythresh(I) method (MATLAB) computes a global threshold, level, that can be used to convert an intensity image to a binary image with im2bw(). The graythresh function uses Otsu's method, which chooses the threshold to minimize the intraclass variance of the black and white pixels

d) **Watershed segmentation:** The final binary image is subjected to marker controlled watershed segmentation. Any grayscale image can be viewed as a topographic surface where high intensity denotes peaks and hills while low intensity denotes valleys. You start filling every isolated valley (local minima) with different colored water (labels). As the water rises, depending on the peaks (gradients) nearby, water from different valleys, obviously with different colors will start to merge. To avoid that, you build barriers in the locations where water merges. You continue the work of filling water and building barriers until all the peaks are under water. Then the barriers you created gives you the segmentation result.

e) **Finding the largest connected component:** Bwconncomp(BW) method (MATLAB) returns the connected components found in the binary image BW. bwconncomp uses a default connectivity of 8 for two dimensions, 26 for three dimensions, and conndef(ndims(BW),'maximal') for higher dimensions.

f) **Final image:** The binary image with largest component is used as mask on the original image for the segmentation of cytoplasm from it which is our proposed cytoplasm.

ii) **Nucleus Extraction Algorithm:** This segment of algorithm is composed of following stages a) K-Mean clustering b) Converting clustered image to binary image c) Finding the largest connected component d) Final image. Figure II show a stepwise execution of Nucleus extraction algorithm processing an image outputted form cytoplasm extraction algorithm.

a) **K-Mean clustering:** The extracted cytoplasm color image is subjected to K-mean clustering for the classification of pixels into different classes. Due to diversity of colors in the extracted image of cytoplasm number of segments taken are six.

b) **Converting Clustered image to Binary image:** The image after clustering is converted into gray scale image and then binary image by finding regional minima. imregionalmin(I) method (MATLAB) returns the binary image BW that identifies the regional minima in I. Regional minima are connected components of pixels with a constant intensity value, and whose external boundary pixels all have a higher value. In BW, pixels that are set to 1 identify regional minima; all other pixels are set to 0.

CERVICAL CELL SEGMENTATION TOOL

c) **Finding the largest connected component:** Bwconncomp (BW) method (MATLAB) returns the connected components found in the binary image BW. bwconncomp uses a default connectivity of 8 for two dimensions, 26 for three dimensions, and conndef(ndims(BW),'maximal') for higher dimensions.

d) **Final image:** The binary image with largest component is used as mask on the extracted image of cytoplasm from cytoplasm extraction algorithm, for the segmentation of nucleus which is our proposed nucleus.

Testing

The data set consists of cervical cell image which has been collected from research done by Sarwar, Abid, et al (2016)[15] from the department of pathology, of Govt. Medical College Jammu, Sher-I-Kashmir Institute of Medical Sciences (SKIMS), Srinagar, Acharya Shri Chander College Of Medical Sciences, Jammu. The data set consist of 2500 images out of which we worked on 200 images. Our tool successfully segments the image with an accuracy of approximately 86%.

Figure I i) Cytoplasm Extraction Algorithm

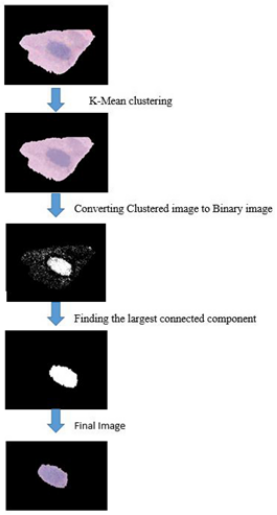


Figure II ii) Nucleus Extraction Algorithm

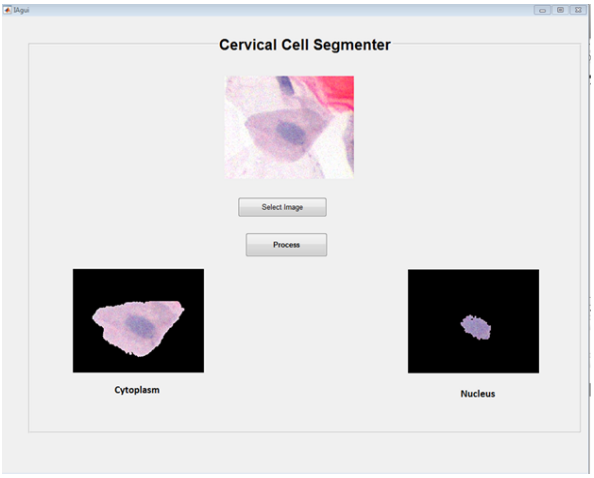


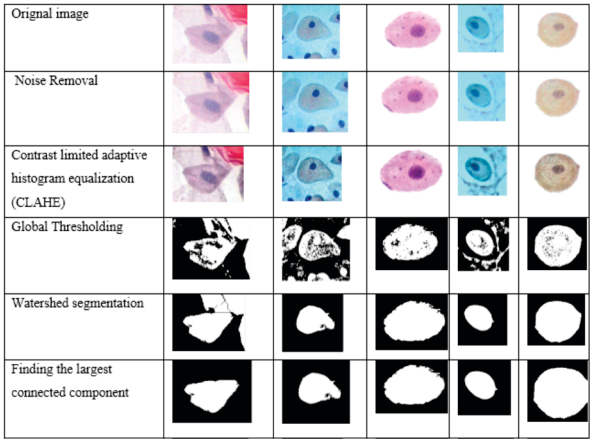
Figure III

Interface of Cervical Cell Segmenter

Figure III shows interface of Cervical cell segmenter in which we can select an image using select button. After selection the image got displayed in top most boxes. Once the image got displayed we can then click the process button. The processing will take the selected image and extract the cytoplasm and nucleus from it. The extracted cytoplasm and nucleus are displayed separately in two labeled boxes.

Below given are some tested images

Figure IV shows stepwise processing of proposed algorithm for five images taken from the database



Extracted cytoplsam and applying K-Mean clustering					
Converting clustered image to binary image					
Finding the largest connected component					
Final image of extracted nucleus					

CONCLUSION

Cervical cancer is one of the leading causes of cancer death in females worldwide. The disease can be cured if the patient is diagnosed in the pre-cancerous lesion stage or earlier. A major problem in the automation of cervical cytology screening is the segmentation of cell images. Our work completely focuses on segmentation of nucleus and cytoplasm of cervical cells as it plays a pivotal role in diagnosis. The data set used for processing consists of 200 images which have been taken from different medical colleges of Jammu and Kashmir. While working on our dataset we come across variety of images. Some images contain multiple cells some with overlapped cytoplasm. Our tool uses the algorithm which divides the problem into two sub parts i) cytoplasm extraction ii) nucleus extraction. The cytoplasm extraction algorithm finds the cytoplasm with accuracy of approximately 86% and nucleus extraction algorithm finds the nucleus with accuracy of approximately 89%.

Future Scope

The tool developed so far has an average accuracy of approximately 87% and is tested only for our data set. Hereafter we intended to refine our algorithm by increasing its accuracy and test it on different data sets. The aim is to build a generalized algorithm which works on different data set with an increased efficiency.

REFERENCES

[1] Chaturvedi, A., & Gillison, M. L. (2010). Human papillomavirus and head and neck cancer. In Epidemiology, pathogenesis, and prevention of head and neck cancer (pp. 87-116). Springer, New York, NY.

[2] Muñoz, N., Bosch, F. X., De Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K. V., ... & Meijer, C. J. (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. New England journal of medicine, 348(6), 518-527.

[3] World Health Organization. Fact sheet no. 297: cancer. February 2006. Retrieved 2007-12-01.

[4] GLOBOCAN 2002 database: summary table by cancer. Archived from the original on 2008-06-16.

[5] Shidham, V. B., Mehrotra, R., Varsegi, G., D'Amore, K. L., Hunt, B., & Narayan, R. (2011). p16INK4a immunocytochemistry on cell blocks as an adjunct to cervical cytology: Potential reflex testing on specially prepared cell blocks from residual liquid-based cytology specimens. Cytojournal, 8.

[6] Kent A. HPV vaccination and testing. Rev ObstetrGynecol 2010;3(1):33e4. PMC 2876324. PMID 20508781.

[7] Human Papillomavirus (HPV) and Cervical cancer: Fact sheet, WHO, 2016.

[8] Li, K., Lu, Z., Liu, W., & Yin, J. (2012). Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. Pattern recognition, 45(4), 1255-1264.

[9] Jusman, Y., Ng, S. C., Osman, A., & Azuan, N. (2014). Intelligent screening systems for cervical cancer. The Scientific World Journal, 2014.

[10] Fu, K. S., & Mui, J. K. (1981). A survey on image segmentation. Pattern recognition, 13(1), 3-16.

[11] Rodenacker, K., & Bengtsson, E. (2003). A feature set for cytometry on digitized microscopic images. Analytical Cellular Pathology, 25(1), 1-36.

[12] Melo, J. C., Cavalcanti, G. D. C., & Guimaraes, K. S. (2003, July). PCA feature extraction for protein structure prediction. In Proceedings of the International Joint Conference on Neural Networks, 2003. (Vol. 4, pp. 2952-2957). IEEE.

[13] Zongker, D., & Jain, A. (1996, August). Algorithms for feature selection: An evaluation. In Proceedings of 13th International Conference on Pattern Recognition (Vol. 2, pp. 18-22). IEEE.

[14] Sarwar, A., Sheikh, A. A., & Koul, N. Segmentation of Cervical Cells for Automated Screening of Cervical Cancer: A Review.

[15] Sarwar, A., Suri, J., Ali, M., & Sharma, V. (2016). Novel benchmark database of digitized and calibrated cervical cells for artificial intelligence based screening of cervical cancer. Journal of Ambient Intelligence and Humanized Computing, 7(4), 593-606.