Pap Smear Screening Test and it's Applications in Cervical Cancer Detection and Classification

S.Athinarayanan
Assistant Professor / Dept of Information Technology
PSN College of Engineering and Technology
Melathediyoor, Tirunelveli District, Tamilnadu, India
aadhinarayan23@gmail.com

Dr.M.V.Srinath
Director / Dept of MCA
STET Women's College,
Mannargudi, Sundarakottai,
Tamilnadu, India

Abstract: A large group of women population, in the world is affected by the cervical cancer in their different age groups. Hence, most of the researchers, pathologists and also more number of collegiate have provided more number of solutions, to identify this cancer from the test images of pap smear screening test. But their results represent, only to an extent, whether the patient is affected by cancer or not, and do not specially identify the exact nature and severity of the cancer. In this paper, it is proposed, to identify the cancer and classify the stage of the cancer from the pap smear test images with high accuracy. The proposed work is based on segmentation, enhancement, and feature extraction and classification process. Test results show the mean and area value based classification of the normal and abnormal cell and also to identify the stage of the cervical cancer. Hence the results would be more useful for the pathologists to reduce their work load and minimize the human error while improving the accuracy of diagnosis.

Keywords: cervical cancer, segmentation, enhancement, nuclei detection, feature extraction, classification.

I. INTRODUCTION

Cervical cancer is the most common malignancy among the women in developing countries. Cervical cancer develops over a prolonged period covering two to three decades [3]. Cervical cancer is the most common form of cancer in women under 35 years of age and the second most commonly occurring cancer in women of all ages, worldwide [1].

Cervical cancer is a preventable disease, and, unlike most other cancers, it can be easily detected by a routine screening test. Currently, cervical smear screening is the most popular method used to detect the presence of abnormal cells developing in the cervix. The method is as follows: Using a small brush, a cotton wad stick, or a wooden stick, a specimen is taken from the uterine cervix, smeared onto a thin rectangular glass plate a slide, and colored, making it easier to examine the cells under a microscope. The purpose of the smear screening is to diagnose premalignant cell changes before they progress to a cancer.

Pap test is the most popular and effective screening test for cervical cancer. By extracting and observing morphology of cervical cells, doctor will classify cervical cells based on Bethesda system; normal, *Low Grade Squamous Intraepithelial Lesion* (LSIL) or *High Grade Squamous Intraepithelial Lesion* (HSIL) cell [4].

Automation of diagnosis of cancer cell image has been studied for a few decades in the field of medical image processing technology. Most of studies involved development of diagnosis support system that solved problems of pathologist [11]. Many techniques of image processing were proposed to handle the problems such as cell segmentation and classification in the development of these systems. Each nucleus was segmented from its surrounding cytoplasm, using a series of automated fast morphological transforms with octagonal structuring elements [12].

Contrast is one of the factors that influence the accuracy of interpretation of diseases, based on medical images by pathologists or radiologists. To date, contrast enhancement process plays an important role in enhancing the quality and contrast of medical images [8]. Prior studies have proven that, contrast enhancement techniques are capable of removing unwanted noises and enhance the brightness and contrast of medical images [5][9].

Contrast is one of the factors that influence the accuracy of interpretation of diseases, based on medical images by pathologists or radiologists. To date, contrast enhancement process plays an important role in enhancing the quality and contrast of medical images [8]. Prior studies have proven that, contrast enhancement techniques are capable of removing unwanted noises and enhance the brightness and contrast of medical images [5][9].

The detection and segmentation of the nuclei in images containing cell overlapping and cell clusters has been successfully addressed by several studies [7]. Classification of Pap smear images are based on the calculation of features extracted from the areas of the nucleus and the cytoplasm [6]. These features are usually based on shape and intensity characteristics of the objects of interest.

Basically a cancer cell has large nucleus area which means, it contains more severity of the lesion like low grade or high grade squamous intraepithelial lesion. Based on this percent of lesion that occurs in the nucleus, the cancer stage could be classified under the given four stages. That is 1. CIN 0 (Normal), 2. CIN 1(Mild dysplasia) 3. CIN 2(Moderate dysplasia) and 4. CIN 3 (Severe dysplasia).

The aim of this work is to develop a cervical cancer detection system that is able to detect and categorize cervical

cells into normal and cancerous cells based on mean and area value of the nucleus in an image and also to classify the given cell into the stage of the cancer.

II. METHODOLOGY

The proposed methodology and the working process of the proposed method are explained briefly under the following sections.

A. Flow Chart:

The following Flow chart shown in Figure 1 represents the work flow of how to detect the cervical cancer and how to analyze the stage of the cervical cancer from the test images of Pap smear screening test images.
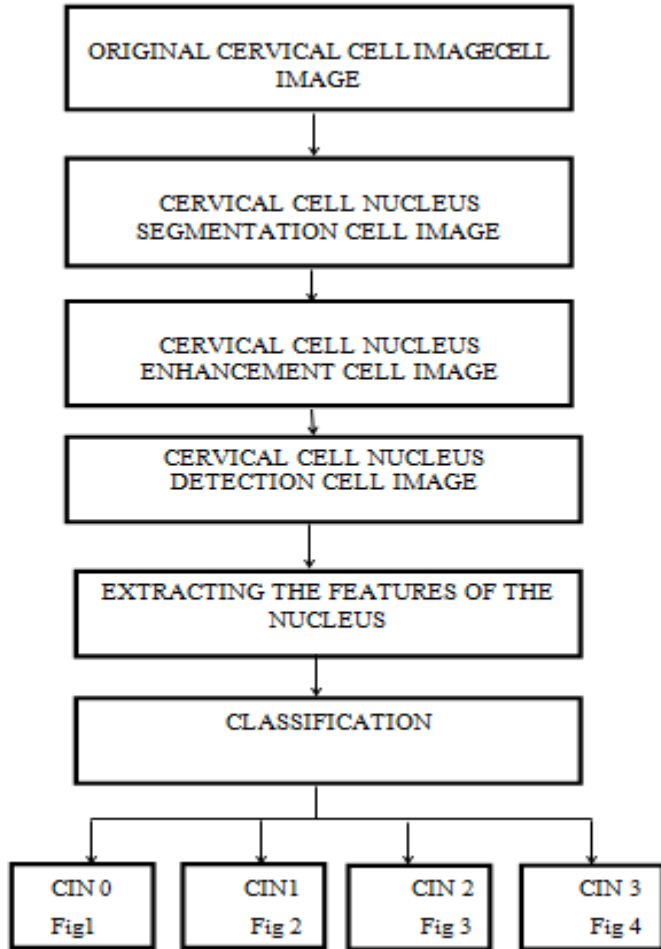
Figure: 1 1.CIN 0 (Normal dysplasia), 2.CIN 1 (Mild dysplasia), 3. CIN 2 (Moderate dysplasia) and 4.CIN 3 (Severe dysplasia) of the cervical cancer.

B. Nucleus Segmentation:

The Original thinprepcervical cell color Image, whether it is a normal, bright or dark image, first it is converted into their respective Red, Green and Blue plane Images.

Then each R, G and B plane of an original image was first locally threshold by using given equations 1, 2 and 3 that is the mean value of the original input image was multiplied with the factors α, β and γ, by adjusting the value of this factors we get incomplete segmentation of the nucleus. So Cytoplasmic

regions were removed by performing a Morphological closing operation using a structure element smaller than the smallest nucleus and nuclear in homogeneity was corrected by an Morphological opening of similar size. The resulting segmented image for the respective R, G and B plane were used for identifying the nucleus with more details than the original image and it is given in equations 4,5 and 6.

$$MI_{R=} MEAN(I_R) * \alpha \qquad (1)$$

$$MI_{G=} MEAN(I_G) * \beta \qquad (2)$$

$$MI_{B=} MEAN(I_B) * \gamma \qquad (3)$$

$$SI_R = ((TI_R \bullet MC_R) \circ MO_R), \qquad (4)$$

$$SI_G = ((TI_G \bullet MC_G) \circ MO_G), \qquad (5)$$

$$SI_B = ((TI_B \bullet MC_B) \circ MO_B), \qquad (6)$$

Where SI_R, SI_G and SI_B are the resulting segmented image for the respective R, G and B plane. TI_R, TI_G and TI_B is the threshold image for the respective R,G and B plane, MC_R, MC_G and MC_B MO_R, MO_G and MO_B are structuring elements for the respective R,G and B plane. The symbols $\bullet$ and o denote morphological closing and opening, respectively.

Then by combining R, G and B Channel Segmented image SI_R, SI_G, SI_B we get the segmented cervical cell image as in equation 7.

$$SI_{RGB} = ((TI_{RGB} \bullet MC_{RGB}) \circ MO_{RGB}) \qquad (7)$$

Where SI_{RGB} is the resulting segmented color image, TI_{RGB} is the threshold image; MC_{RGB} and MO_{RGB} are Structure elements of both the closing and opening operation.

C. Nucleus Enhancement:

After the segmentation process is completed, we get the segmented nucleus from its surrounding cytoplasm. But the results were somewhat light portioning of the nucleus. For this reason, authors have gone for the enhancement process to enhance or increase the contrast of the nucleus. In this enhancement process based on the two equations 8 and 9, it is used only after the segmentation process is completed.

The two equations are based on the normal, bright or dark image of the segmentation output image. That is, if the image was normal means we use the equation 8 and if it is bright or dark means we use the equation 9.

$$E_{RGB}(i,j)= [(SI_{RGB}(i,j) * SI_{RGB}(i,j)) / (MAXRGB - MINRGB)] \qquad (8)$$

$$E_{RGB}(i,j)=[(SI_{RGB}(i,j)*SI_{RGB}(i,j))/(MAXRGB)] \qquad (9)$$

Where

$E_{RGB}(i,j)$: The Enhanced RGB value of the pixel
$SI_{RGB}(i,j)$: The segmented RGB value of the Pixel
MAX_{RGB}: Maximum RGB value of the segmented Image
MIN_{RGB}: Minimum RGB value of the segmented Image

Finally both the equations have successfully increased the contrast of the nucleus in the segmented image and also the nucleus details are very clearly seen.

D. Nucleus Detection:

After the enhancement process is completed, we have increased contrast of the nucleus in an image. But if in that image if we need to see the details of the nucleus very clearly, we have to identify the boundary or edge of the nucleus or shape of the nucleus in a very clear manner as required.

For this reason authors go for nucleus detection process. In this process authors just apply a line scanning algorithm to detect the edges of all the cervical cell of the enhanced image. Before applying line scanning algorithm, we need to convert each one of the R, G, B plane of the enhanced image into binary image.

After converting to the binary image all the line scanning directions namely horizontal, vertical, +45 degree and -45 degree each and every one of the R, G and B plane of the binary images are applied. Then by combining all the directions of the line scanning output to get the edge of the nucleus in the binary image is done. But the results represented that all the edges of the nucleus and it's color was white and remain pixels of an image represent black color.

From these final output of binary edge image, R, G and B Plane of the edge or boundary or shape of the images, authors come to a conclusion that the green plane of an image can be used to find the exact edge of the nucleus and also apply these edges into the green plane of the enhanced image by using the following two conditions.

a. If the pixel value of the green plane of an edge image is 1 means, there is no need to change the corresponding pixel value of the green plane of enhanced image.

b. If the pixel value of the green plane of an edge image is 0 means, the corresponding pixel value of the green plane of enhanced image can be taken into the account.

Then, to view the exact shape of the nucleus in the color image, in addition to the green plane of the enhanced image by using the above two conditions, we must take the red and blue plane of the enhanced image.

The final image, thus arrived, will be clearer to detect the nucleus with the edges than the original image.

E. Feature Extraction:

After detecting the nucleus, we must extract the features of the nucleus to find the growth of the nucleus that is normal or abnormal because then only we can identify whether the given cell is affected or not affected by the cancer.

For this reason nucleus feature extraction process is done. In this process, First of all the converted images, the color image of nucleus detection is converted into the gray scale image. Then, some of the features like area, mean, perimeter, standard deviation of the nucleus are extracted by using XROI command in IDL Language.

The use of this command helps us to find out the feature values of the nucleus, from the gray scale image.
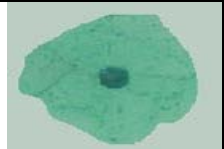
F. Classification:

After the feature of the nucleus is extracted, it is further moved into the classification process.
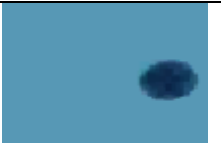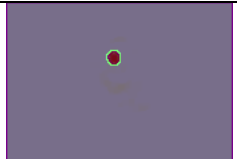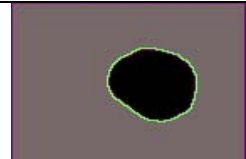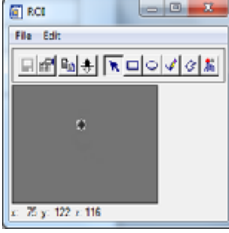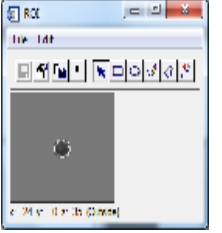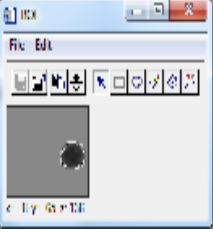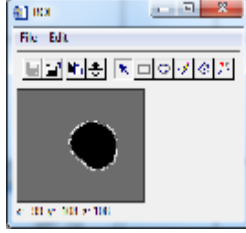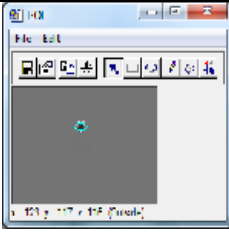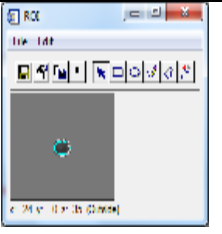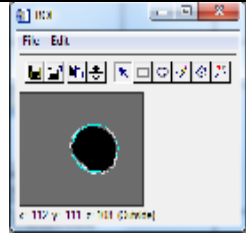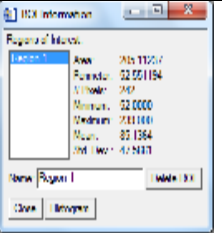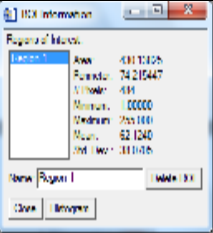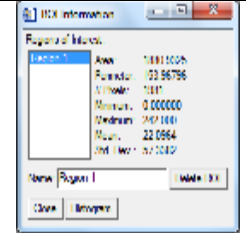
The classification process is mainly focused on, two main features of the nucleus like mean value and area value. By using this value we can identify whether the given cell is affected or not affected by cancer based on normal or abnormal growth of the nucleus. Then these results help us to find whether further the given cell can be classified into the category of given four stages, namely 1.CIN 0 (Normal dysplasia), 2.CIN 1 (Mild dysplasia), 3.CIN 2 (Moderate dysplasia) and 4.CIN 3 (Severe dysplasia) of the cervical cancer.

III. RESULTS AND DISCUSSIONS

In this section a detailed explanation of digital image processing techniques with the corresponding image results were discussed based on the following two tables namely Table.1for detecting and classifying the stage of the cancer corresponding CIN 0, CIN 1, CIN 2 and CIN 3.

Table 1: Image Results of the Cancer Stages

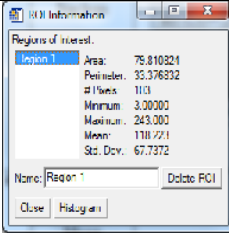| S. No	Image Names and Important Factors	Images and Resulting Factors Results			
		Fig. 1	Fig. 2	Fig. 3	Fig. 4
a)	Input Image				
b)	Segmentation				
	α	0.57	0.50	0.45	0.57
	β	0.57	0.77	0.70	0.57
	γ	0.77	0.77	0.77	0.77

c)	**Enhancement**				
	Normal				**Normal**
	Bright or **Dark**	Bright	Dark	Bright	
d)	**Detection**				
e)	**Feature Extraction**				
	ROI Window				
	Selection Of Nucleus				
	ROI Information				
f)	**Classification**				
	Mean	118.223	85.1364	62.1240	22.0964
	Area	79.810824	205.11237	430.13825	1880.9325
	Stage of the Cancer	CIN 0	CIN 1	CIN 2	CIN 3

In above Table.1,the Fig. 1,2,3 and 4 (a) represents the original input images and by applying the eqn 7 we get the nucleus segmented image as shown in Fig.1,2,3 and 4 (b) with adjusting threshold factor α, β and γ, from this image to increase the contrast of the nucleus by applying the eqn. 8(normal image) or 9(bright or dark image), the enhancement result shown in Fig.1,2,3 and 4 (c).Then the shape or edge or boundary of the nucleus detected from the resultant enhancement image as shown in Fig.1,2,3 and 4 (d). After the edge detected image the result of features of the nucleus extracted by using the XROI command in IDL language from the ROI window, and the draw an ellipse tool based selected the nucleus of the ROI window and also the values of the features contains ROI information windows are shown in given Fig.1,2,3 and 4 (e). Finally the features of the value based classification of the cell results shown in the Fig.1,2,3 and 4 (f) by using the given Table 2.

Table 2: Range of the Mean and Area values of Cancer Stages

S.No	Mean Range	Area Range	Stage of the Cancer
1	above 100	below 150	CIN 0
2	Above 80 – below 100	above 150 and below 400	CIN 1
3	Above 50 – below 80	above 400 and below 600	CIN 2
4	below 50	above 600	CIN 3

Basically, the cancer detection is made based on decreased mean value and increased area value of the Region Of Interest (ROI).

As compared with the previous researches conducted by Eko Supriyato and Jeremiah, color intensity level of normal cell ranges between [122-150] and [151 – 200] and cancerous cell ranges between [80-100] and [51 – 100] – [2] and also these results were used only for the particular cell was affected or not affected by the cancer, but the present new research will be very useful to the pathologist for not only to identify the cancer but also to classify the given cell into which stage of the cancer by using the mean and area value of the table 2. But the table results were only applicable for single cervical cell that was present in an image not for multiple numbers of cervical cell images. And also sometimes (it's a very rare case), if in the table values specifically mean value is changed for any of the stage, then, we must take the decision based on the area value of the corresponding stage. Then the ranges of the mean and area values of the nucleus are based on the results of tests on more than 100 numbers of samples of all the stages of CIN Images.

IV. CONCLUSION

This work shows that many number of image processing techniques like segmentation, enhancement, detection, feature extraction and classification are used for finding the details of the region of interest from the initial cervical cell and are seen very clearly with more accuracy than the original image. Further extracting some of the features like mean and area of the region of interest for finding the given cell is normal or abnormal is carried out. Then based on the results, the cell might be classified into the following four stages of the cancer. That is 1.CIN 0 (Normal dysplasia), 2.CIN 1 (Mild dysplasia), 3. CIN 2 (Moderate dysplasia) and 4.CIN 3 (Severe dysplasia) of the cervical cancer. Hence the results reported will be very useful to the pathologists for their further investigations.

V. REFERENCES

[1] Beral.V and M. Booth, "Prediction of cervicalcancer incidence and mortality in England andWales," Lancet, vol. 1, p. 495, 1986

[2] Eko supriyanto, Nur azureen M. Pista, lukman hakim ismail,bustanur rosidi, tati latifah mengko: Automatic Detection System of Cervical Cancer Cells Using Color Intensity Classification. ISBN: 978-1-61804-019-0. Recent research in computer science. (2011).

[3] Kiviat, N. (1996). Natural History of CervicalNeoplasia: Overview and Update. Am J. ofObstet Gynecol. 175. 1099-1104.

[4] Kurman.R.J, D. E. Henson, A. L. Herbst, K. L. Noller, and M. H.Schiffman, "Interim guidelines of management of abnormal cervical cytology,"J. Amer. Med. Assoc., vol. 271, pp. 1866–1869, 1994

[5] Lim, E. E., Venkatachalam, P. A., Ngah, U. K. & Khalid, N. E. A. (1999). Liver Disease Diagnosis by RegionGrowing. Proceedings of International Conference onRobotics, Vision and Parallel Processing for Automation. 1.38-45

[6] Marinakis, Y., Marinaki, M., Dounias, G.: Particle swarm optimizaton for Pap-smear diagnosis.Expert Systems with Applications 35, 1645–1656 (2008)

[7] Marinakis, Y., Dounias, G., Jantzen, J.: Pap smear diagnosis using a hybrid intelligent schemefocusing on genetic algorithm based feature selection and nearest neighbour classification.Computers in Biology and Medicine 39, 69–78 (2009).

[8] Mat Isa, N.A., Mashor, M.Y. & Othman, N.H.(2003). Contrast Enhancement ImageProcessing Technique on Segmented PapSmear Cytology Images. Proc. of Int. Conf.on Robotics, Vision, Information and Signal Processing. 118-125.

[9] Ngah, U. K., Ooi, T. H., Sulaiman, S. N. &Venkatachalam, P. A. (2002). Embedded Enhancement Image Processing Techniques on A Demarcated Seed Based Grown Region. Proc. of Kuala Lumpur Int. Conf. on Biomedical Engineering. 170-172.

[10] Rafael, C. Gonzalez and Richard E. Woods."Digital Image Processing," 2nd edition,Prentice Hall, 2002.

[11] Tanaka, T., Joke, T., Oka, T. (2001). "CellNucleus Segmentation of Skin Tumor UsingImage Processing", IEEE Proc. of 23rd AnnualInt. Conf. Engineering in Medicine andBiology Society, Istanbul, Turkey, Vol.3, pp.2716-2719

[12] Yong H. Lee. Algorithms for mathematicalmorphological operations with flat topstructuring elements. SPIE Applications ofDigital Image Processing, pages 33-45, 1982.