Papsmear Image based Detection of Cervical Cancer

Sreedevi M T
Dept. of ECE
RNS Institute of Technology
Bangalore-61

Usha B S
Dept. of ECE
RNS Institute of Technology
Bangalore-61

Sandya S
Dept. of ECE
RNS Institute of Technology
Bangalore-61

ABSTRACT

In this paper, a newapproach is proposed for the early detection of cervical cancer using Papsmear images. Regular Papsmear screening is the most successful attempt of medical science and practice for the early detection of cervical cancer. Manual analysis of the cervical cells is time consuming, laborious and error prone. This paper presents an algorithm for classifying Cervical cells as normal or abnormal. It is tested on 80Papsmear images and the experimental results show that the algorithm is on par with the results obtained by earlier work and gives satisfactory results in terms of sensitivity (100%) and specificity (90%).

General Terms

Cervical Cancer, early detection, mass screening programme,Papsmear

Keywords

Segmentation, Feature Extraction, Classification, Sensitivity, Specificity

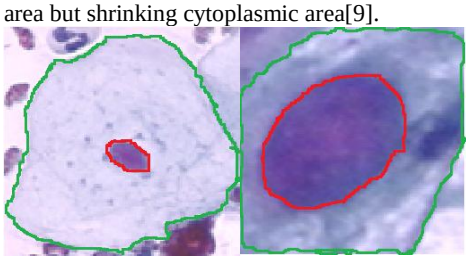
1. INTRODUCTION

Cervical cancer is one of the most common cancers among women worldwide. Itcauses loss of productive life in women both due to early death as well as prolonged disability[1].The primary reason is lack of awareness of the disease and access to screening and health services. Government of India has undertaken several cancer control programme, but these measures have not been effective to reach rural regions due to issues related to trained manpower, infrastructure, logistics, quality assurance and frequency of screening involved which has resulted in increase of mortality rate[2].According to the survey, one woman dies every seven minutes of cervical cancer and by 2025, it is estimated to be one death in every 4.6 minutes [3].There has been a regular campaign against cervical cancer for the last 30 years in India, but this has had little impact on the morbidity and mortality from the disease, with India ranking fourth worldwide [4].

Cervical cancer if detected early has very good prognosis.Cervical screening using Papsmear images is one of the most effective ways of detecting and diagnosing the disease even at an early pre-cancerous stage.During mass screening program there will be huge number of samples to be analyzed and diagnosed and the current manual screening methods are time consuming and restricts the capabilities of the cyto-technicians in diagnosing more samples in shorter time. Therefore there is a need for a support system for faster analysis of samples. Work has been carried out by Byriel [5], Martin [6] and Norup [7] for classification of Papsmearimages. Martin and Norup used Champ Software developed by DIMAC Imaging for segmentation followed by classification of cervical cells. The main methods used by Martin were Hard C-means(HCM), Fuzzy C-means(FCM) and Gustafson-Kessel clustering(GK) for classification.Byriel used neuro-fuzzy classification method to classify the cervical cells.Byriel used Fuzzy C-means(FCM), Gustafson-Kessel clustering(GK) and ANFIS(Adaptive Network based Fuzzy Inference System) for classification of cervical cells. Methods based on Watershed transform have also been used for the segmentation of nucleus boundary [8].With reference to the work done by Martin, Byriel and Norup we are proposing a new approach to develop a support system for early detection of Cervical cancer. The work is based on single cell Papsmear images. The database developed by Martin is available in open source for research purpose and is used for analysis and validation.The paper is organized as follows. In section2 description of cervical cancer is given, in section 3, the proposed method for early detection of cervical cancer is discussed, in section4, experimental results and discussions are presented, section5 concludes the paper.

2. OVERVIEW OF CERVICAL CANCER

Cervical cancer is a malignant disease that develops in the cells of the cervix or the neck of the uterus. These cells do not suddenly change into cancer. Instead, the normal cells of the cervix first gradually develop precancerous changes which later turn into cancer. Cancerous cells show increasing nucleus area when compared to normal cells. This characteristic feature can be used to do a first level of classification of the cervical cells as normal or abnormal. The Figure (1a) and (1b) shows normal cell and abnormal cell [6].Normal cell has smaller nucleus area and a very large cytoplasmic area whereas abnormal cell has increased nucleus area but shrinking cytoplasmic area[9].



Figure(1a). Normal Cell **Figure(1b). Abnormal Cell**

Nucleus Region

Cytoplasm Region

3. PROPOSED METHOD

The proposed approach for cervical cancer detection using Papsmear images is as shown in figure (2). The input images used in this method are conventional Papsmear images which are microscopic optical images in .bmp format. In this work we have considered single cell Papsmear image for evaluating the cell as normal/abnormal. The first stage is the preprocessing block which constitutes color conversion and enhancement to prepare the image for further processing. The next stage is the processing block whose objective is to segment the nucleus from a single cell. Following this stage is the feature extraction where the area of the nucleus is extracted for classification of the cervical cells as normal or abnormal. Each stage is discussed in detail in section 3.

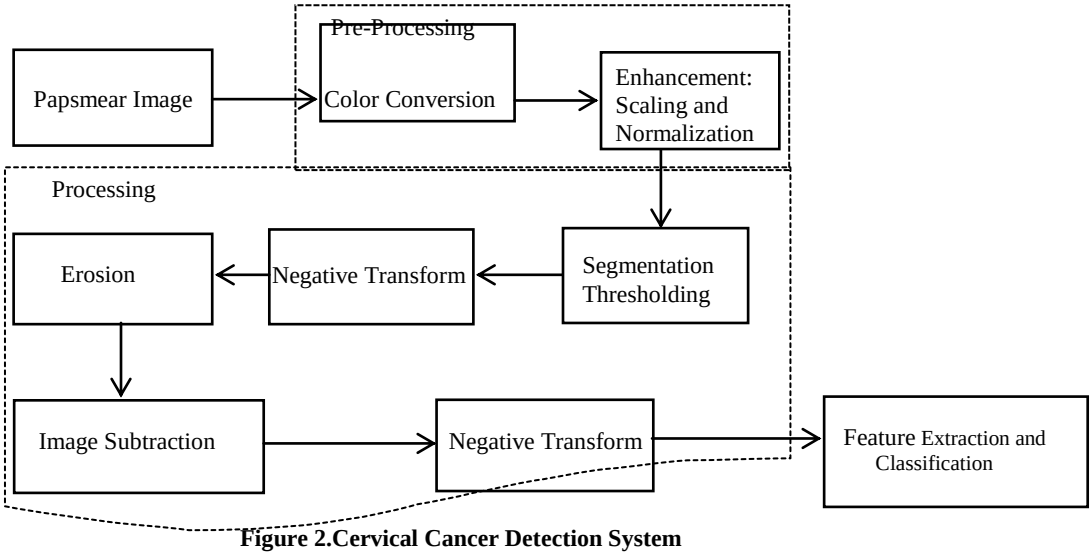


Figure 2.Cervical Cancer Detection System

3.1 Pre-processing

The pre-processing stage prepares the image for further processing, analysis and interpretation. ThePapsmear images are coloured optical images which are of poor quality due to stains used to colour the cells and uneven lighting across the field of view. The colour information is insignificant compared to the intensity of the image. Hence in the pre-processing stage, the first step is to convert the RGB image to Intensity/Gray scale image. RGB image is converted to grayscale by forming a weighted sum of the R, G, and B components using the equation:

$$Igray = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

It is required to have a well-defined boundary between the nucleus and the cytoplasm to extract nucleus from the cell. The contrast of the Papsmear images is poor and the lack of homogeneity in image intensity makes further processing difficult [11]. Hence there is a need to improve the contrast of the image so as to get clear edges and boundaries of nucleus necessary for further processing.The accuracy of image segmentation and feature extraction mainly depends on this factor. The next block is the enhancement block and linear scaling is used to improve the contrast of the image so that only the nucleus boundary is clearly seen from the background image making it more suitable for segmentation. The grayscale image is scaled by a scaling factor α given by the equation:

$$Iscale = \alpha * Igray \quad (2)$$

Where α is determined empirically.

Linear enhancement results in image values either suppressed or enhanced beyond the range of [0,255]. Normalization is performed on the enhanced image to confine the pixel values in the range of 0 to 255. This image is the input to the processing block to extract the nucleus.

3.2 Processing

3.2.1 Segmentation

The first process in this stage is segmentation which is the fundamental step in image analysis or image understanding. It aids in extracting regions of interest in an image. A good segmentation must be able to separate objects from the background to obtain the region of interest [12]. Image regions are expected to have homogeneous characteristics (e.g., gray level, or colour), indicating that they belong to the same object or are facets of an object, implying the possibility of effective segmentation. In this paper an automatic bi-level thresholding method is used which uses the global mean of the image to calculate the threshold value iteratively. The Iterative thresholding method is used to segment the nucleus from the Papsmear image. The key parameter in the thresholding process is the choice of the threshold value. In this method, the threshold value T is iteratively calculated and the pixels with gray values less than T are classified as object pixels, and pixels with gray values greater than T are classified as background pixels.

The threshold value is iteratively updated using the following steps:

Step1: The input is scaled imageIscale [Figure (3c)]. The initial threshold value (T) is selected by calculating the average of the maximum and minimum pixel value of the image.

$$Tinitial = (Z0 + Z1)/2$$

Z0 = maximum pixel value of Iscale

Z1 = minimum pixel value of Iscale

Step2: Initialize S0, S1, N0, N1 to zero where S0, S1 are sum of pixel values above the threshold and below the threshold

value respectively. N0, N1 are the count of pixels above and below the threshold respectively.

Step3: If Iscale(i,j)>$Tinitial$ then update S0 and N0 as

$$S0 = S0 + Iscale(i,j)$$
$$N0 = N0 + 1$$

Else, update S1 and N1 as

$$S1 = S1 + Iscale(i,j)$$
$$N1 = N1 + 1$$

Step4: Calculate$T0 = S0/N0$

$$T1 = S1/N1$$

Step5: $TTnew = (T0 + T1)/2$

Step6: Update new threshold value as

$$T = TTnew$$

Step7: Repeat until$Tinitial \neq TTnew$

Step8: Perform segmentation using the new threshold value

If Iscale(i,j) > T, make the pixel values to 1 else 0.

3.2.2 Morphological Operation

The negative of the segmented image is more suitable as it aids in removing unwanted clusters using morphological erosion faster than the true image [Figure (3e)]. The resultant image has nucleus region along with small spurious objects which can be removed by performing erosion operation using the equation 3 [10]

$$A \ominus B = \{x : B + x \subseteq A\} \quad (3)$$

The eroded image is subtracted from the negative image to get contours of the segmented objects as shown in figure (3g).After erosion the resultant image consists of nucleus region along with other regions. It is required to select only the nucleus region by some criteria. From experimentation, it is observed that the nucleus is a bigger region than the other regions. Hence the next step is the labeling the objects followed by selecting the biggest cluster of pixels among the various clusters. This selects only the nucleus region as shown in figure(3i).

3.3 Feature Extraction

Transforming the input data into the set of features is called feature extraction. It involves algorithms to detect and isolate various desired portions, parameters or shapes (features) of a digitized image. The features of the nucleus such as area and centroid are extracted using region properties.The area of the nucleus is the number of pixels inside the nucleus. In the Training Phase the area of the nucleus of the normal cells and abnormal cells is calculated. The mean area of the nucleus of normal cells(□norm) and abnormal cells(□abnorm) are calculated using the equation 4.

$$Mean\ nucleus\ area = \frac{\sum_{i=0}^{n} Nucleus\ Area}{n} \quad (4)$$

Where n = number of images

3.4 Classification

The classification of the image is based on the areaof the nucleus of cells. Feature extraction results showed that normal cells have smaller nucleus with arealess than the calculated threshold value and abnormal cells have larger nucleus with areagreater than the threshold value. Hence classification of cells into normal and abnormal cells is done based on area parameter of the nucleus. In the Testing Phase the area of the nucleus of the input image will be calculated and compared with the mean threshold nucleus area of normal cells(□Tnorm) calculated. If the nucleus area of the input image is greater than □Tnorm then it is classified as abnormal else it is classified as normal.

3.5 Performance measures

In medical diagnostics, sensitivity and specificity are the performance measures which specify the accuracy of diagnosis method. It is desired to have a very high (100%) sensitivity and specificity of any diagnostic method. Sensitivity is the ability of a test to correctly identify those with the disease (true positive rate), whereas specificity is the ability of the test to correctly identify those without the disease (true negative rate). The sensitivity test gives the probability of a positive test given that the patient is ill. This can also be written as:

$$Sensitivity = \frac{Number\ of\ TP}{Number\ of\ TP + Number\ of\ FN} \quad (5)$$

The specificity test gives the probability of the negative test given that the patient is well.

$$Specificity = \frac{Number\ of\ TN}{Number\ of\ TN + Number\ of\ FP} \quad (6)$$

- True positive(TP): Sick people correctly diagnosed as sick
- False positive (FP): Healthy people incorrectlydiagnosed as sick
- True negative(TN): Healthy people correctly diagnosed as healthy
- False negative(FN): Sick people incorrectly diagnosed as healthy.

4. EXPERIMENTAL RESULTS

The experimentation is done using 80 samples of Papsmear images, out of which 40 (20 normal and 20 abnormal) single cell images are used for Training Phase and 40 images for Testing Phase. The results of training phaseand testing phase are shown in Table 1, Table 2, Table 3 and Table 4.

Table 1: Results of Training Phase: Normal Cells

Images	Area	Centroid [i ,j]
Pap1	1030	144.53 , 150.78
Pap2	899	114.10,87.96
Pap3	2517	179.75 , 157.57
Pap4	1289	136.21 , 95.65
Pap5	1443	151.90 , 142.54
Pap6	764	130.09,134.8
Pap7	1473	119.54 , 121.44
Pap8	630	130.70 , 103.40
Pap9	933	134.43 , 169.56
Pap10	1631	142.42 , 102.63
Pap11	1765	147.04 , 101.05
Pap12	506	158.75 , 155.77
Pap13	1231	131.03 , 137.13
Pap14	1067	118.31 , 160.93
Pap15	1134	132.72 , 121.35
Pap16	296	105.61 , 139.84
Pap17	971	126.12 , 142.69
Pap18	210	182.2 , 68.0
Pap19	721	135 , 112.55
Pap20	1188	148.60,152.82
Mean area of the nucleus = 1135		

Table 2: Results of Training Phase: Abnormal Cells

Images	Area	Centroid [i ,j]
A1	19939	135.78,136.44
A2	16286	123.54,143.00
A3	17759	98.89,99.69
A4	48091	138.47,133.62
A5	15035	115.01,86.58
A6	27623	122.61,102.80
A7	13074	103.57,124.63
A8	28684	151.98,131.88
A9	19109	121.87,133.59
A10	25106	143.33,142.67
A11	13086	86.86,118.33
A12	20411	114.06,121.88
A13	25329	138.17,131.13
A14	19795	114.59,134.71
A15	27769	159.65,143.04
A16	13922	136.33,110.95
A17	20191	100.82,141.49
A18	16725	111.63,138.19
A19	18392	110.15,130.67
A20	23663	149.58,116.65
Mean area of the nucleus = 21500		

Table 3: Results of Testing Phase: Normal Cells

Images	Area	Centroid [i ,j]
Pap21	200	137.64,119.15
Pap22	246	143.29,139.19
Pap23	409	139.64,99.73
Pap24	263	108.72,115.86
Pap25	212	148.39,112.12
Pap26	257	133.70,145.94
Pap27	204	121.31,152.24
Pap28	276	120.30,138.59
Pap29	290	104.70,148.45
Pap30	**1813**	**146.17,130.07**
Pap31	330	157.44,118.16
Pap32	322	142.51,119.65
Pap33	196	141.43,117.70
Pap34	118	155.53,181.41
Pap35	402	124.12,137.17
Pap36	**1772**	**154.92,148.59**
Pap37	439	128.79,122.56
Pap38	261	139.93,143.45
Pap39	150	148.26,112.32
Pap40	240	138.72,143.31

Table 4: Results of Testing Phase: Abnormal Cells

Images	Area	Centroid [i ,j]
A21	8660	183.77,107.05
A22	12145	194.31,174.61
A23	20628	119.92,183.30
A24	30784	139.31,114.95
A25	29516	161.55,119.87
A26	27519	158.95,132.29
A27	14765	166.79,136.50
A28	16883	138.32,119.33
A29	14855	136.16,110.85
A30	17761	111.21,136.75
A31	13754	155.20,182.32
A32	18059	110.36,130.93
A33	16233	155.85,147.89
A34	13518	130.60,124.24
A35	19215	126.75,145.63
A36	17985	156.23,186.27
A37	21563	165.38,176.45
A38	15746	116.89,154.72
A39	18234	137.12,145.86
A40	17962	142.75,136.25

From the results of the Training Phase it is observed that the mean value of nucleus area for 20 normal cells is norm= 1135 pixels and that for 20 abnormal cells is abnorm= 21500 pixels.There is a clear differentiation between the nucleus area value of normal and abnormal cells and hence classification of the cells can be done based on area parameter. In this work we have considered the mean threshold value to be Tnorm= norm + Tn (where Tn=500 pixels).

In the Testing Phase, the proposed method is tested with 40 Papsmear images (20 normal and 20 abnormal images).The area of the nucleus of the input image was calculated and compared with the mean threshold nucleus area (Tnorm =1635). The cell was classified as normal if nucleus area of the input image was lesser than Tnorm and classified as abnormal otherwise. Based on this method of classification a sensitivity of 100% and specificity of 90% was achieved. The results are validated with the results of Martin [6]. The proposed method classified all the 20 abnormal cells correctly as abnormal as classified by Martin[6] achieving sensitivity of 100% and among 20 normal cells, the proposed method could correctly classify 18 cells as normal and incorrectly classify 2 cells as abnormal (Image No. Pap30 and Pap36 highlighted in Table 3) achieving specificity of 90%. As it is desired to have high specificity, improvement can be made by including otherparameters for classification like nucleus to cytoplasm ratio, perimeter, intensity etc.The overall error rate as considered in the work done by Martin [6] is given by the equation:

$$Overall\ Error\ Rate = \frac{FN + FP}{TP + FN + TN + FP} * 100\% \qquad (7)$$

- True positive(TP): Abnormal cells correctly identified as abnormal
- False positive (FP): Normal cells incorrectly identified as abnormal

- True negative(TN): Normal cells correctly identified as normal

- False negative(FN): Abnormal cells incorrectly identified as normal.

The overall error rate achieved by Martin [6] with Supervised Gustafson-Kessel clustering(GK) is 6.06%. The proposed work achieves an overall error rate of 5% with 40 Papsmear images used in the testing phase. Hence the proposed method gives better classification of cells with acceptable error rate.

The figure (3) shows the segmented results of nucleus of normal and abnormal cells.

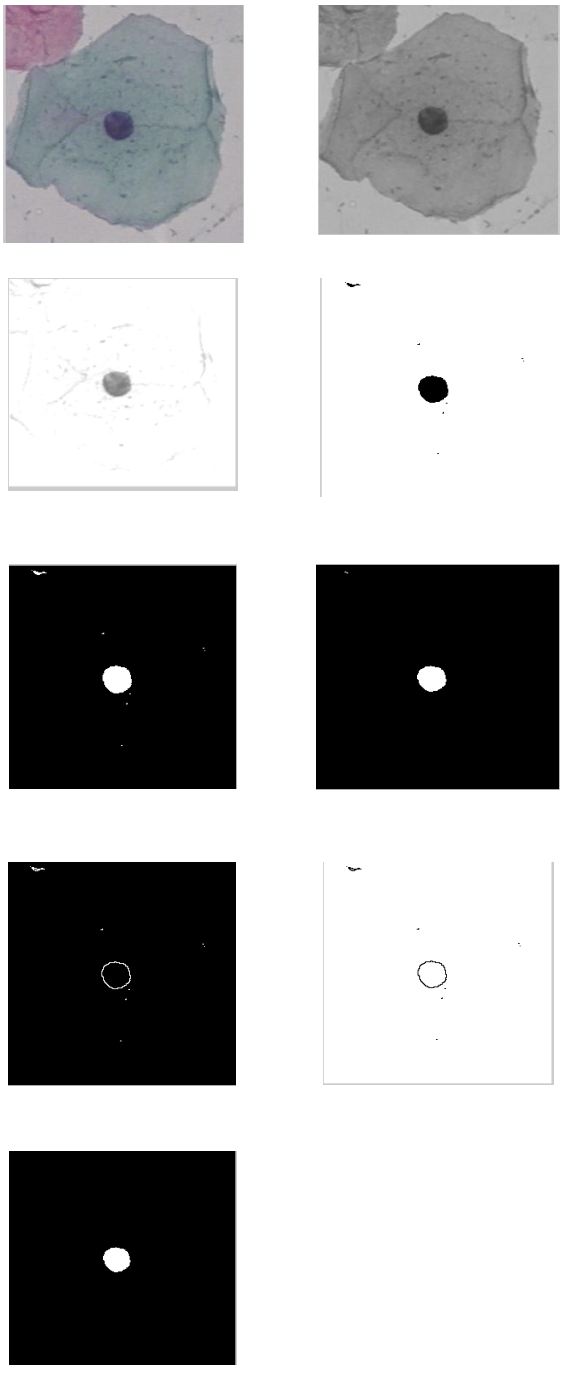


Figure 3(a). Segmented Results of Normal Nucleus

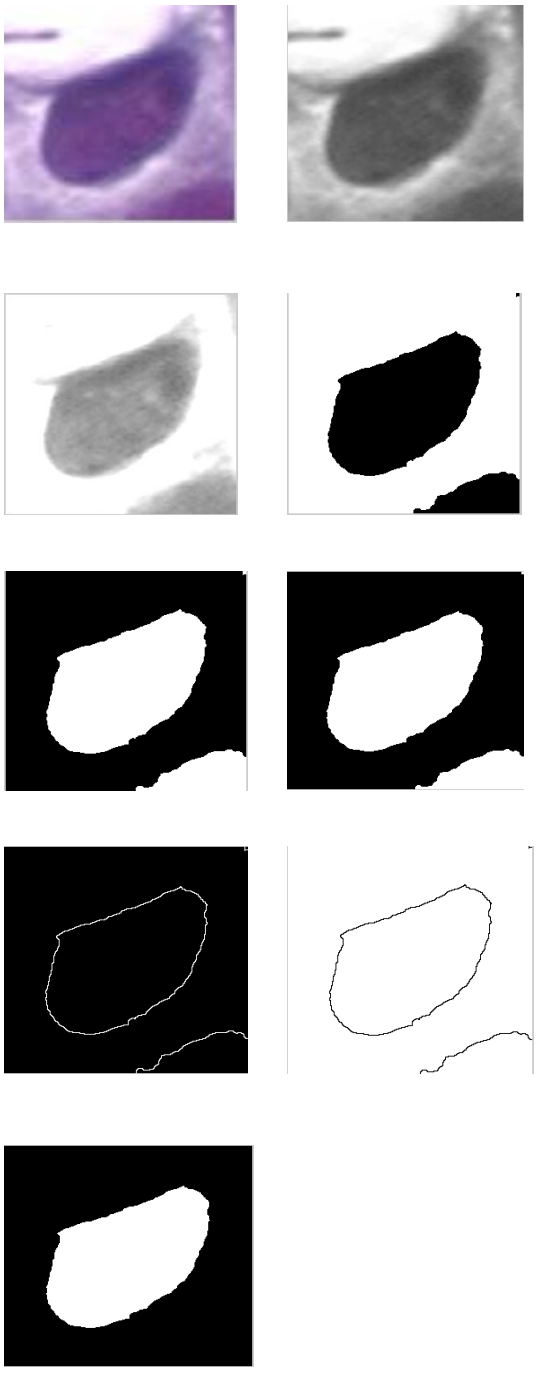


Figure 3(b). Segmented Results of Abnormal Nucleus

a	b
c	d
e	f
g	h
i	

(a) Original image, (b) Gray image, (c) Scaled image, (d) Segmented image, (e) Negative transformed image, (f) Eroded image, (g) Subtracted image, (h) Negative transformed image, (i) Nucleus extraction

5. CONCLUSION

In this paper, we have proposed an approach for classification of cervical cell as normal or abnormal using area of the nucleus as a feature. The experimental results show that this method gives good classification and achieves sensitivity of 100% and specificity of 90% with acceptable overall error rate of 5%. The classification results are validated with the benchmark database prepared by Martin[6]. The proposed methodserves as a basis for first level classification of Papsmear images for detection of cervical abnormality using area of the nucleus as the parameter.

6. ACKNOWLEDGMENTS

Our thanks to RNS Institute of Technology for the Lab support provided for executing the work. Our extended thanks to Dr.Vipula Singh, Professor, Department of ECE, RNSIT for valuable technical inputs.

7. REFERENCES

[1]Cervical Cancer in India, South Asia Centre for chronic disease website available at:

http://sancd.org/uploads/pdf/cervical_cancer.pdf

[2]Dinshaw KA, Shastri SS, PatilSS,"Cancer Control Programme In India: Challenges For The New Millennium", Health Administrator Vol: XVII, Number 1: 10-13,pg.

[3]The Times of India newspaper website available at:

http://articles.timesofindia.indiatimes.com/2012-02-04/mumbai/31024533_1_cervical-cancer-breast-cancer-globocan

[4]Cervical Cancer-Overview and Incidence website available at:
http://www.medindia.net/patients/patientinfo/cervicalcancer-incidence.htm#ixzz1oDusUgV1

[5] Jens Byriel,"Neuro-Fuzzy Classification of Cells in Cervical Smears", M.Sc Thesis

[6] Erik Martin," Pap-Smear Classification" Thesis Report, 22nd September 2003

[7]Norup, "Classification of pap-smear data by transductive

Neuro-fuzzy methods", Master Thesis

[8] Marina E. Plissiti, Christophoros Nikou1 and Antonia Charchanti," Combining shape, texture and intensity features for cell nuclei extraction in pap smear images", Pattern Recognition Letters, Vol.32, No.6, pp.838-853, 2011

[9]BustanurRosidi, NorainiJalil, Nur. M. Pista, Lukman H. Ismail, EkoSupriyantoTati L. Mengko "Classification of Cervical Cells Based on Labeled Colour Intensity Distribution" International Journal of Biology and Biomedical Engineering, Issue 4, Volume 5, 2011

[10]J.Serra "Image Analysis and Mathematical Morphology", Academic Press, London, 1982

[11]M.E. Plissiti, A. Charchanti, O. Krikoni and D.I. Fotiadis, "Automated segmentation of cell nuclei in PAP smear images"

[12]Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Pearson Education, Inc. and Dorling Kindersley Publications, Inc.