GREEDY APPROACH IN ANALYZING MULTI-CLUSTERED CELL NUCLEI ON PAP SMEAR IMAGES

Peter Kulandai Raj[1] Dr. S. Sheeja[2]

ABSTRACT

The Advent of Cytological driven cancerdiagnosis widely accepted with the constanteffort of Dr.Papincolaou's research worksince 1943. Demanding need to automate thisimage processing approach for precancerousdetection has gone through multiple phases inthe last decade [1] [2]. This paper proposes aGreedy Algorithm approach for an automatedprocessing of the Pap smear images inidentifying the malignant cells using chromaticbehavior of the stained images. The uniquecontribution of this approach is to process theimages with the less number of cycles anddiscriminate normal cells from cervical cancercells thereby increase the efficiency of thesystem while not overlooking the accuracy. Ascompared to the classical approaches appliedfor the same medical challenge 0.07% to0.03% false positive result is proven by thisapproach. This paper also highlights the needfor overlapped nucleus processing [3] usingthe same approach that helps in detecting thehidden cells during automated image analysis.There has been similar work carried out bymany forerunners but this paper uniquely address how chromatic behavior of the imagecan be used in identifying cell differences.

[1]Research Scholar, Dept of Computer Science, Karpagam University. Email : faffyroypeter@gmail.com

[2]Associate Professor and Head, Dept of Computer Applications, Karpagam University. Email : sheejaajize@gmail.com

Keywords : Dysplasia, Segmentation

I. INTRODUCTION

Ever since the establishment of cell theory in the beginning 19th century, which recognized the cell as the fundamental building unit of life, biologists have sought to detail the underlying principles. Significant discoveries were made over the course of many decades of research [4]. The longing desire to attain a complete understanding of the cellular mechanism and how to manipulate them is the constant effort spent till this day by many researchers across the globe. Image Segmentation is often considered as the keystone of image analysis process. Specifically since cellular morphology is an important phenotypic feature that is symptomatic of the physiological state of a cell, and the since the cell contour is often required for subsequent analysis of intracellular process (Zooming in or zooming out) the problem of cell segmentation has received increasing attention in the past [4]

The father of microbiology Antonie Philips van Leeuwenhoek (1632-1723) first paved a starting point in 1670 AD towards improving and exploiting microscopic imaging for studying life at the cellular level. The initial use of computers for the cell analysis date back more than 50 Years. For the first time around 1950s Computer systems were developed to automate the classification of smears of exfoliated

cells with an objective to detect cervical cancer applying threshold-based decision rules to serial one-dimensional microscopic line scans of a specimen [5]. 1960's automated processing of 2D images were experimented to find out the differential counting of White Blood Cells (WBC - Leukocytes) based on simple chromatic and morphological measurements [6], in mid-1970's routine clinical test hit the market with computer circuits enabling paralyzing the analysis of image of the previous cell while grabbing the image of the previous cell and at the same time locating next cell in the specimen.

II. Background

The Papanico laou smear method, is a medical procedure to find pre-cancerous cells in the uterine cervix. A small cytological specimen from the uterine cervix (see figure 2) is collected with a special cyto-brush and smeared onto a glass slide. Then the slide is stained using the Papanicolaou method, so the different components of the cells are emphasized with specific colors. This glass slide is then viewed under a microscope, so cyto-technicians can diagnose the cells on the glass slide to identify the intensity of the pre-cancerous cells.

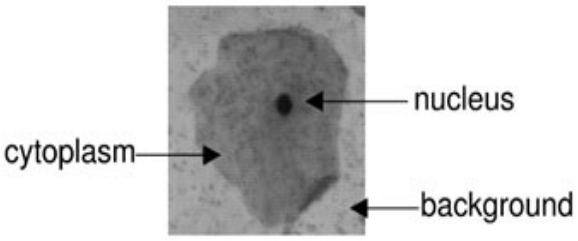


Figure 1. Single Pap smear Image

Cyto-technicians use several different features to get a cell diagnosis. The size, color, shape and the texture of the nucleus and cytoplasm is used. The density of cells in a certain area, can influence the diagnosis. It takes a skilled cytotechnician, to be able to differentiate between the different cells. [7] Every glass slide, can contain up to 300.000 cells. Therefore it is a time consuming job viewing the slides

The Papanicolaou cells

Ideally specimens are taken from several areas of the cervix. Depending on the area, the cyto-brush, cotton stick or the wooden stick is used. The specimens most often contain cells from the columnar epithelium and the squamous epithelium. The columnar epithelium is located in the upper part of the cervix, and the squamous epithelium in the lower part. In between these two areas, the metaplastic epithelium is found, also called the transformation zone

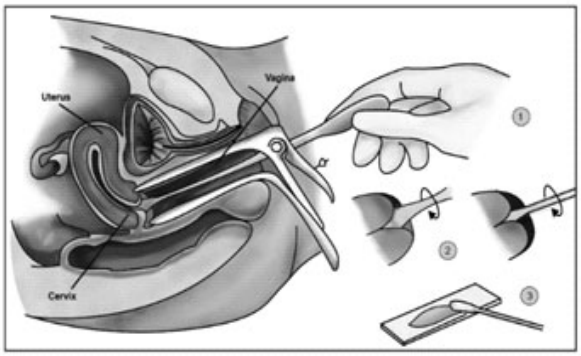


Figure 2: Extraction of Vaginal Cells (1.Extract 2. by rotating the swab the cells is collection, 3. Staining the cells with reagents on the glass for microscopic observation)

In the squamous epithelium there are 4 layers of cells. The cells start outbeing formed at the basal layer and while maturing they move out through theparabasal layer, the intermediate layer and at last out through the superficiallayer. The cells in the basal layer divide and deliver cells to the layers above it.

While the cells mature and move through the layers, [8] they change shape, colorand other characteristics. When the cells reach through the superficial layerthey are rejected and replaced by the cells coming from below. The basal layerhas small round cells with a relative big nucleus and a little cytoplasma. Whenmaturing, the nucleus becomes smaller and the cytoplasma becomes bigger.The shape of the cells become less round the more mature they are.The columnar epithelium only contains a single layer of cells, the basal layer here contains columnar cells and reserve cells. The reserve cellsdivides into new reserve cells and into columnar cells.The metaplastic epithelium consists of reserve cells from the columnar epithelium. When the cells have matured fully in the metaplastic epithelium,they look like the cells found in the squamous epithelium.When the genetic information in a cell somehow has changed, the cell willnot divide as it should. This is a pre-cancerous cell [9].

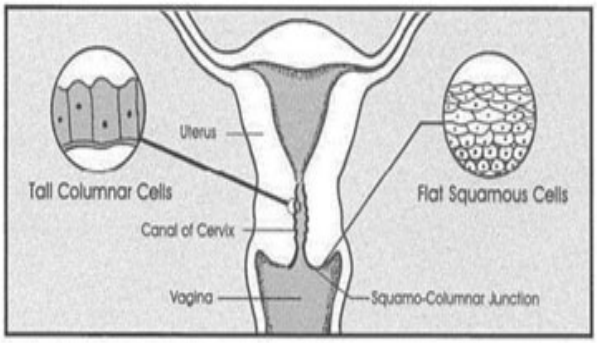


Figure 3 : The uterus in details and the location of a) columnar cells and b) squamous cells

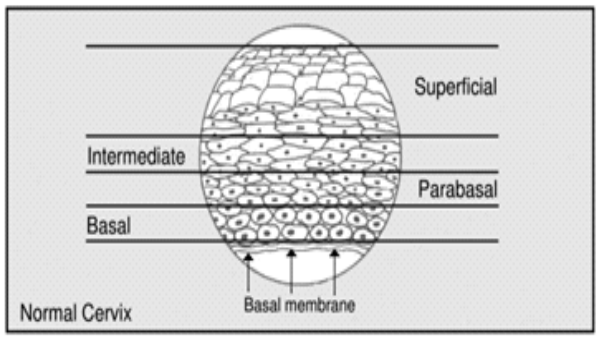


Figure 4 : Development of the squamous cells through the four layers

Depending on whichkind of cell that is dividing incorrectly, it is given diagnoses like dysplasia andcarcinoma in situ. The dysplastic cells are divided up into mild, moderateand severe dysplastic. The graduation into different degrees of dysplasia aredetermined from the probability of the cells later on turning into malignantcancer cells. A high amount of the mild dysplastic cells will disappear without becoming malignant, whereas severe dysplastic cells

quite likely will turninto malignant cells. In medical terms these are divided into 2 different main diagnoses :

1. Dysplasia. The term "*plasia"* means growth, and dysplasia means disordered growth.The cervical dysplasia are normally divided into 3 types: *mild, moderate and severe*, describing the risk that the cells turn into malignant cancer cells. [10] Mild means of course lowestrisk. The characteristics of cells in dysplasia depends on the kind. In the mild dysplasiathey have enlarged and light nucleus. For the moderate dysplasia the nucleus is larger anddarker. The nucleus has begun to deteriorate, which is seen as a granulation of the nucleus.In severe dysplasia the nucleus is large, dark and often deformed. The cytoplasm is darkand small when compared to the nuclei.

2. Carcinoma-in-situ. Carcinoma-in-situ means "*cancer in place"* and is characterized ofvery large nucleus. In the past, there was a tendency to treat "carcinoma-in-situ" as a muchmore serious problem than severe dysplasia, when in fact they are essentially the same The pre-malignant cells are characterized by alarger nucleus and a bigger N/C ratio [11][12]

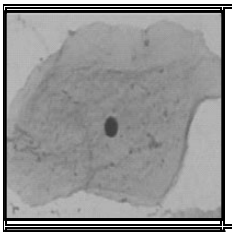
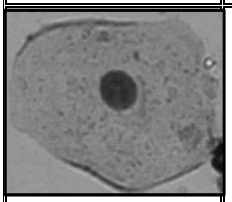
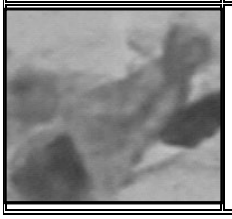
The N/C ratio is given by :

III. Classification

Complexity in cervical cytological image analysis is that of getting more false positive results due to lack of classifying the false cases as opposed to true cases and also it is observed that it is highly complicated process to generalize the cell pattern in identifying the exact shape of the carcinoma cells during automation. However based on the Human experience this has been categorized before automated processing. As defined below there are two major categories by which any cells that is subjected to smear image analysis will be categorized. [13] Cells that do not fall under any of the category is considered as wrong image or out of the scope images. Those images will not be providing the results though the right approach is followed during image processing methods and techniques.

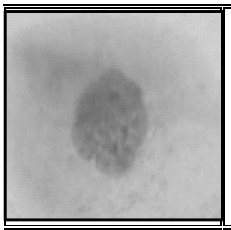
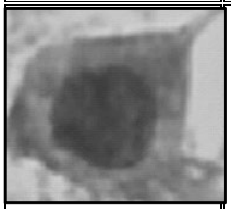
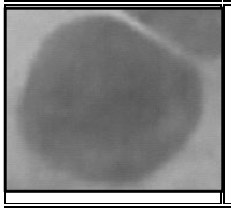
Normal Cells

Table 1: Normal Cells Classification

	Classifier01 **Superficialsquamous** N-Shape: Flat /Oval Nucleus : Very Small N/C : Very Small
	Classifier 02 **Intermediate squamous** N-Shape: Round Nucleus : Nucleus Large N/C : Small
	Classifier 03 **Columnar** N-Shape: Column Like Nucleus : Large N/C : Medium

Abnormal Cells

Table 2 : Abnormal Cell Classification

	Classifier04 **Mild dysplasia** Nucleus :light / large N/C : medium
	Classifier 05 **Moderate dysplasia** Nucleus : Large / dark Cytoplasm : dark N/C : Large
	Classifier 06 **Severe dysplasia** Nucleus : Large /dark/ deform Cytoplasm : dark N/C : Very Large
	Classifier : 07 **Carcinoma in Situ** Nucleus : Large /dark/ deform N/C : Very Large

The above [14] [15]seven cell classification is oriented on the core attributes such as Size, area, shape and brightness of both nucleus (N) and cytoplasm (C). Listed below are the parameters considered for the processing of the stained cell image on the segmented area

Table 3 : Parameters for cervical cancer cells detection

Nucleus	Cytoplasm
area (B)	area (C)
brightness (E)	brightness (F)
shortest diameter(G)	shortest diameter (K)
longest diameter (H)	longest diameter (L)
elongation (I)	elongation (M)
roundness (J)	roundness (N)
perimeter (O)	perimeter (P)
maxima (R)	maxima in C (T)
minima (S)	minima in C (U)
N/C ratio (D)	
relative position (Q)	

Parameter Definition for C N

- **Area** - Calculated by counting the corresponding pixels of the segmented picture, pixel area $(0.201\mu m)^2$
- **Brightness**–it is calculated as the average perceived brightness, that is the function of the colors wavelength where Y= 0.299 * Redμ + 0.587 * Greenμ + 0.114 * Blueμ(Average intensity of the colors)
- **Longest Diameter** – The biggest distance between any of the two points of the border
- **Shortest Diameter** – The shortest distance between any of the two points of the border

- **Elongation** – ration between shortest and longest distance ,

  $N_{elong} = N_{short}/N_{long,}$ $C_{elong} = C_{short}/C_{long}$

- **Roundness** – ratio between the actual area and the area inside the circle given by the longest diameter of the object

  $N_{circle} = \delta /4 * N^2_{long}$

  $=> N_{roundedness} = N_{ared}/N_{circle}$

  $C_{circle} = \delta /4 * C^2_{long}$

  $=> C_{roundedness} = C_{ared}/C_{circle}$

- **Perimeter** – The length of the perimeter around the object
- **Nucleus Relative Position**- a measure of how well the nucleus is centered in the cytoplasm. It is calculated by finding the distance between the nucleus center and the center of the cytoplasm
- **Maxima / Minima** – Count of how many pixel values of maximum / minimum value inside of three pixels
- **N/C ratio** – denotes how small the nucleus area is compared to the area of the cytoplasm

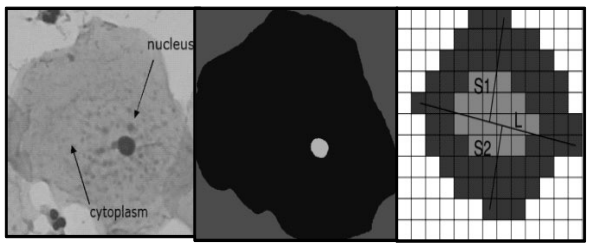


Figure 5 : a) Class 01 normal Picture b) segmented picture of class 01 c) a binary cell picture with background (white) cytoplasm (black) and Nucleus (grey)

IV. Experiment

Data Collection

The images are collected from microscope on 400X magnification with the resolution of 568*768 in 24 bit colors scaling 0.201 μm/pixel from the Government Hospital Chennai and Holy Spirit Hospital Mumbai, and image standards and preprocessing were followed from existing processed image database. These images were captured by the microscope connected through camera and PC.Taking into consideration of the above the parameters a native image processing tool is developed in identifying or isolating the nucleus by which the 'Carcinoma in situ' are identified.

Preprocessing and Training

Every image collected from the healthcare center is classified and preprocessed with the cytotechnologist expert's advice. Images that cannot be processed were removed from the collected images and thus a total of 355 images were collected who fall into 1, 2, 3 classes and 445 images of 4, 5, 6, 7 classes. The images of class 1, 2, 3 were taken into the system by training them for the models that corresponds to Non-cancerous smears conditions. 4, 5, 6, 7 images were taken to train the system for identifying cancerous smears. Every image is of single nucleus. Every image is uniquely stored in the database for identification with its class types. These images are used as a benchmarked source for identifying cancerous positive or negative

Process Flow

Patient's Pap smear image is loaded to **RES**ult **PRO**vider(RESPRO 2.0 – The software specifically designed to compare other cervical cancer detection algorithms against Greedy Algorithm)

1) Pap smear image registration happens with RESPRO

2) Pap smear image filters are applied so as to make it processable

3) The 19 factors around Cytoplasm and Nucleus are identified

4) Based on the parameters True Positive (4, 5, 6, 7) or False Positive (classes of 1, 2, 3) are identified and categorized

5) Every nuclei along with cytoplasm is marked in with respective colors so that marking is visible for users

6) Depending upon the Expert Knowledge and the benchmark result is provided on the image

Successive Greedy Algorithm

Classical algorithms referred under Figure 7 depicts most of the algorithms used in image cervical smear image analysis.These are some of the famous approaches followed by various methods in order for the same purpose. The bigger challenge with these algorithm is always they take in to account of the image as only true positive. This increases the error rate. Images that has lot of defects will still return the false positive results due to the factor it is taking for processing the image. However this Greedy Algorithm proposed here stands unique in identifying the distance of every nuclei and cytoplasm by identifying its unique points in terms of its curve at its pixel level.The important advantage of processing at every pixel level will validate the image first and then start accepting the image for the positive result. Let G be the graph that is identified for processing, there exists multiple nodes that has combinations of nucleus (N), cytoplasm (C) using grey level region identification methods. Load the image in to the system, convert the picture into its underlying image array, the array contains the every pixel as per the image and a three dimensional array that has X, Y, Z where X contains x axis of the pixel and Y contains Y axis of the pixel and Z contains the color code that is associated at the pixel background

Core Algorithm

Convert Image into basic RGB for all pixels in Images

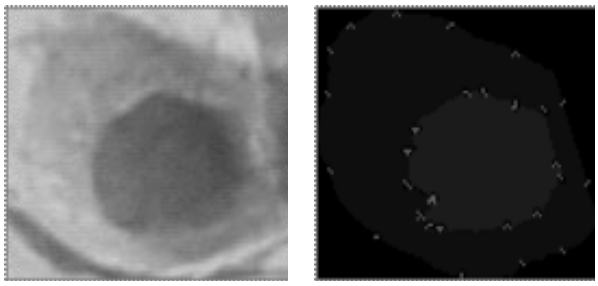
1. Read pixel G(x,y)

2. Identify the chromatic (R |G|B)

3. From the database compare theR|G|B benchmark that identifies a class type

4. If the cell color falls under definedranges 0 to 255 Mark them as possible region of Cell Nucleolus

4.1. Mark them with labels End Condition End of all pixels Collate all Labels and draw lines betweenevery label



a b

Figure 6 :

a) Cancerous cells before processing,

b) Cancerous cells is identified with its axis marked with boundaries

For every G (N | C) identify the points using the masking color applied on the region along with the row matches such as Black, Blue, Deep Blue. Whenever the pixel matches a black and subsequently with the Blue color the points are marked thus the image identifies its outer boundary at the first iteration going through every pixel of the image. In the second iteration the array is compared against the Blue color with the Deep Blue Color, since the image is going to have only these three colors thus system process the image on the whole and marks the points for its every edges. Once all the e is identified the standard parameters (Refer Table 3) value are calculated from its pixels.In the last iteration the marks that were separated against Black, Blue and Deep Blue are marked. [16] [17-24] The standard parameters such as shape, size of the Nucleus along with the Cytoplasm is taken into account and basing on the classifier 4,5,6,7 the classification is still narrowed down to dysplasia or Carcinoma in situ is identified. Deviation of accuracy in images were discarded those that are not proper for image processing those image are neglected

Table 4 :

Acuteness identification table for the cancer cells

Degree of Dysplasia			
Normal	**Mild**	**Moderate**	**Severe**
Nucleus Area			
20-50	50	50	50
Nucleus Intensity			
Dark	light	Dark	dark
Cytoplasm Intensity			
Light	light	Dark	dark
N/C Ratio			
1-2%	10-20%	20-50%	> 50%

V. Results And Discussion

In this paper we have highlighted about how theMultiple regions finding algorithms in Pap smear images have been extensively covered in the history of Pap smear image analysis on the whole, the comparison of the existing algorithms against the proposed Greedy Algorithm has been compared across 50 Pap smear mages of which, the results provided by Greedy algorithm.

Table 5 : Indicative of the accuracy as compared to classical approaches

Method	Images	Dysplasia Classes	Deviation %	Accuracy %
Unsupervised GK Cluster Method	500	1+2+3Vs. 4+5+6+7	90 - 10%	**88.66%**
Supervised GK Cluster Method	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**95.56%**
Feature Selection and Unsupervised GK Cluster Method	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**90.08%**
Feature Selection and Supervised GK Cluster Method	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**97.11%**
Unsupervised FCM (Fuzzy C-Means)	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**96.69%**
Supervised FCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**96.94%**
Feature Selection & Unsupervised FCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**98.19%**
Feature Selection & Supervised FCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**98.36%**
Unsupervised HCM (Hard C-Means)	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**94.01%**
Supervised HCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**95.97%**
Feature selection Unsupervised HCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**96.12%**
Feature Selection Supervised HCM	500	1+2+3 Vs. 4+5+6+7	90 - 10%	**97.20%**
Unsupervised Greedy Algorithm	**50**	1+2+3 Vs. 4+5+6+7	**95 - 05%**	**96.02%**
Supervised Greedy Algorithm	**50**	1+2+3 Vs. 4+5+6+7	**95 - 05%**	**98.02%**

As compared to the error rate generated by the other algorithm on the same image has been 0.7% to 0.3 %. The false positive rate was high in most of the algorithms listed down. The Greedy Algorithm works best condition with 0.4% error rate. However there are other improvements such as overlapping images and pixel quality those determine the results.

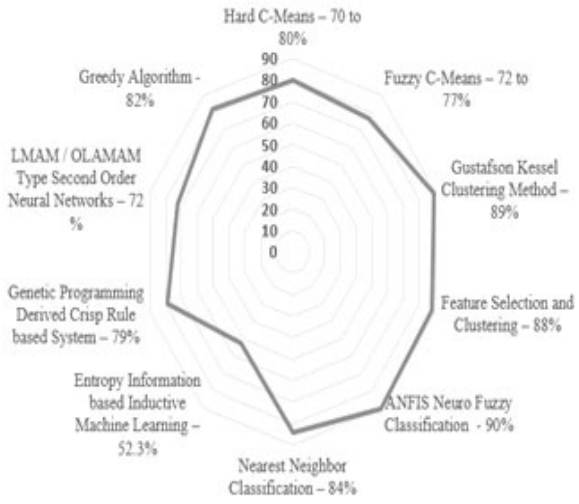


Figure 7 : Efficiency of Greedy Algorithm over conventional approaches and the false positive results percentage is displayed against each approach

VI. Future Scope

Future direction would be to improvise the present Greedy Algorithm at cervical smear image processing so that error rate could be reduced at a great extent lowering the bound from 3% to 0%. The shape of the nuclei and cytoplasm is an important factor in finding the true positives as such this is one of the real time challenge to narrow down to 0%. This algorithm can be used in multi nuclei images especially

when they are overlapped [25]. Because the classification and image processing is carried out at the pixel level using chromatic behavior of the images, [26] proposed GreedyAlgorithm will be a preferred choice in find the hidden nuclei region andcomparing to other algorithms this will yield a good result as the number of steps for converging to the intended result is quicker and shorter

Acknowledgement

Author would like thank Dr. Rajendran of Government Hospital, Chennai and Cytologist Sr. Visalatchi for their expert advice on cervical cancer.

References

[1] Erik Meijering, *"Cell Segmentation: 50 Years Down the Road",* IEEE Signal Processing Magazine, Sep 2012

[2] James S. Duncan, Nicholas Ayache, *"Medical Image Analysis: Progress over Two Decades and the Challenges Ahead",* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, Jan 2000

[3] Marina E. Plissiti and Christophoros Nikou, *"Overlapping Cell Nuclei Segmentation Using a Spatially Adaptive Active Physical Model",* IEEE Transactions on Image Processing, vol. 21, No. 11, Nov 2012

[4] B.Alberts, A.JohnsonJ.Lewis, M. Raff K.Roberts, P.Walter, "*Molecular Biology of the Cell"* 5 th ed. New York, Garland Science, 2007

[5] E. Bengtsson, C.Wählby and J.Lindblad, *"Robus Cell image segmentation methods*" Pattern Recognit. Image Anal. Vol. 14. No. 2 pp. 157-167, May 2004

[6] W. E. Tolles, "*The Cytoanalyzer - An example of physics in medical research*," Trans. NY Acad. Sci., vol. 17, no. 3, pp. 250–256, Jan. 1955

[7] Koss L.: "*The Application of PAPNET to Diagnostic Cytology*". P. Lisboa, E. Ifeachor, P. Szczepaniak (Eds.), Artificial Neural Networks in Biomedicine, (2000)51–68 Springer

[8] Matthews, B.W.: "*Comparison of the predicted and observed secondary structure of T4 phage lysozyme*". Biochemica et Biophysica Acta,405(1975) 442–551

[9] Jens Byriel, "*Neuro-Fuzzy Classification of cells in Cervical Smears*", Feb 1999

[10] Landwehr, D, "*Web based pap-smear classification*", Dept. of Automation, Bldg 326,2800 Lyngby, Denmark, 2006

[11] Erik Martin, "*Pap Smear Classification*" a detailed study pp 3-17 Sep 2003

[12] Gonzalez, R. C. & Woods, R. E. "*Digital image processing*", Pearson Education Inc Publishing Company, 2009

[13] Norup, Jonas,"*Classification of pap-smear data by transductiveneuro-fuzzy methods*" pp 5-25, 2005

[14] Kasabov, N. and Song,Q. "*Denfis: Dynamic, evolving neural fuzzy inference system and its application for time-series prediction*". IEEE Transaction on Fuzzy System,pp 144–154, 2002

[15] Nikolaos Ampazis, George Dounias, and Jan Jantzen, *"Pap-Smear Classification Using Efficient Second Order Neural Network Training Algorithms*", G.A. Vouros and T. Panayiotopoulos (Eds.): SETN 2004, LNAI 3025, pp. 230–245, 2004.Springer-Verlag Berlin Heidelberg 2004

[16] J. M. S. Prewitt and M. L. Mendelsohn, "*The analysis of cell images*," Ann. NY Acad. Sci., vol. 128, no. 3, pp. 1035–1053, Jan. 1966

[17] K. Preston, "*Computer processing of biomedical images*," Computer, vol. 9, no. 5, pp. 54–68, May 1976

[18] Voss, M. Fuzzy Billedbehandling, Technical University of Denmark, Department of Automation, Bldg, 326, Dk2800 Lyngby Denmark, 1996.

[19] NikolaosAmpasis, George Dounias, Jan Jantzen,"*Pap-Smear Classification Using Efficient SecondOrder Neural Network Training Algorithms* ", Springer 2007

[20] Z. Li and K. Najarian, "*Biomedical image segmentation based on shapestability*," inProc. 14th IEEE Int. Conf. Image Process. Sep.–Oct. 2007,pp. 281–284.

[21] M.H.Tsai,Y.K.Chan,Z.Z.Lin,S.F.Yang-Mao,andP.C.Huang,"*Nucleus and cytoplast contour detector of cervical smear image*,"Pattern Recognit. Lett., vol. 29, no. 9, pp. 1441–1453, 2008.

[22] S. F. Yang-Mao, Y. K. Chan, and Y. P. Chu, "*Edge enhancement nucleusand cytoplast contour detector of cervical smear images*,"IEEETrans.Syst., Man, Cybern. B, Cybern., vol. 38, no. 2, pp. 353–366, Apr. 2008.

[23] H.Lin, Y.K.Chanand C.C.Chen, "*Detectionandseg mentation of cervical cell cytoplasm and nucleus*," Int. J. Imag., Syst. Technol.,vol. 19, no. 3, pp. 260–270, 2009.

[24] C.W.Chang, M.Y.Lin, H.J.Harn, Y.C.Harn, C.H.Chen, K.H.Tsai, and C. H. Hwang, "*Automatic segmentation of abnormal cell nuclei frommicroscopic image analysis for cervical cancer screening*," inProc. 3rdIEEE Int. Conf. Nano-Molecular Med. Eng., Oct. 2009, pp. 77–80.

[25] M. E. Plissiti, C. Nikou, and A. Charchanti, "*Automated detection ofcell nuclei in pap smear images using morphological reconstruction andclustering*", IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 2, pp. 233–241, Mar. 2011.

[26] M. E. Plissiti, C. Nikou, and A. Charchanti, "*Combining shape, textureand intensity features for cell nuclei extraction in Pap smear images*",Pattern Recognit. Lett., vol. 32, no. 6, pp. 838–853, 2011

Author's Biography

Mr. Peter Kulandai Raj received his Bachelor and Master of Computer Applications from Indira Gandhi National Open University, Delhi. He is a research scholar doing PhD in Computer Science at Karpagam University. He has been working as technical consultant with Microsoft for the last 5+ years with overall 11 years of extensive industry experience in Dot Net Technologies. His special interest's areas are Medical Image Analysis, Machine Learning and Cloud Computing.

Dr. S.Sheeja, is currently working as Associate Professor in Department of Computer Applications at Karpagam University, Coimbatore. She had completed Ph.D in Computer Science at Bharathiar University in 2015. She has more than 14 years of teaching experience to her credit. Her primary research interests are related to Computer Networks, Mobile Adhoc Networks, Mobile Computing, Image Processing and Data mining. She has published papers in various International and national Journals.