Open access Journal International Journal of Emerging Trends in Science and Technology
IC Value: 76.89 (Index Copernicus) Impact Factor: 4.219 **DOI:** https://dx.doi.org/10.18535/ijetst/v4i3.08

# Cervical Cancer Detection Using Support Vector Machine

Authors
**R.Aruna Sankari, BE[1], S.Bavithra BE[2], Dr J.Nithyashri ME,Ph.D[3]**
[1,2]Dept of Computer Science And Engineering, Jeppiaar SRR Engineering College, Chennai
[3]Associate Professor, Dept of Computer Science And Engineering, Jeppiaar SRR Engineering College, Chennai

**ABSTRACT**
*The pre-cancer can be identified by most efficient method of segmentation and by classification in a pap smear images. The cervical cell cancer can be detected by monitoring exactly the changes in a cell .The major challenging task is to identify the overlapping of cytoplasm which is accurately done in this work. The earlier cervical cancer is diagnosed, the more successfully it can be treated. Regular cervical screening can save thousands of live every year. The majority of these deaths could be prevented if all women had undergone cervical screening. From the global survey say a female should start screening at the age of 21, or within 3 years of her first sexual encounter - whichever occurs first. Cervical screening does not detect cancer, it simply looks for abnormal changes in the cells of the cervix. If left untreated, some abnormal cells can eventually develop into cancer. The input is fetched from the dataset and those images been processed using efficient function of resizing, grayscale conversion and noise reduction. High-level shape information to guide segmentation where cell boundary might be weak or lost due to cell overlapping. In the segmentation Canny edge detection model used to detect the feature. The feature is extracted accurately through histogram of gradient. The entire classification is done by support vector machine. An evaluation carried out using two different datasets ISBI 2015 dataset and SZU dataset of the proposed method over the state-of-the-art methods in terms of classification accuracy.*
***Keywords:*** *cytoplasm, preprocessing; segmentation; extraction; classification; Cervical cancer; overlapping cells splitting; Pap smear screening.*

## INTRODUCTION

Cervical cancer is a cancer arising from the cervix. It is due to the abnormal growth of cells that have the ability to invade or spread to other parts of the body. Human papilloma virus (HPV) infection appears to be involved in the development of more than 90% of cases; most people who have had HPV infections, however, do not develop cervical cancer. Other risk factors include smoking, a weak immune system, birth control pills. Cervical cancer typically develops from precancerous changes over 10 to 20 years. About 90% of cervical cancer cases are squamous cell carcinomas, 10% are adenocarcinoma, and a small number are other types. Diagnosis is typically by cervical screening followed by a biopsy. Medical imaging is then done to determine whether or not the cancer has spread. HPV vaccines protect against between two and seven high-risk strains of this family of viruses and may prevent up to 90% of cervical cancers. As a risk of cancer still exists, guidelines recommend continuing regular Pap smears. Cervical cancer screening using the Pap smear or acetic acid can identify precancerous changes which when treated can prevent the development of cancer. Treatment of cervical cancer may consist of some combination of surgery, chemotherapy, and radiotherapy. Five year survival rates in the United States are 68%. Outcomes, however, depend very much on how early the cancer is detected. Worldwide, cervical cancer is

both the fourth-most common cause of cancer and the fourth-most common cause of death from cancer in women. In 2012, an estimated 528,000 cases of cervical cancer occurred, with 266,000 deaths. This is about 8% of the total cases and total deaths from cancer. About 70% of cervical cancers occur in developing countries. In low-income countries, it is the most common cause of cancer death. In developed countries, the widespread use of cervical screening programs has dramatically reduced rates of cervical cancer. To automate the screening methods, accurately segmenting cervical cells is an essential task, which makes extracting quantitative features of each cell accessible. The structure information of cells (e.g., shape, staining intensity, texture, and number), which is extracted from each individual cell, is vital to the early detection of cervical cancer. At very first step the input is processed through filtering, resizing, grayscale conversion and by noise reduction. After the preprocessing of the image is been segmented by using threshold values. Fuzzy Segmentation is used to find the intensity level. Intensity level detection is based on logical segmentation. By conducting an exhaustive survey of image thresholding methods, categorize them, express their formulas under a uniform notation, and finally carry their performance comparison. The thresholding methods are categorized according to the information they are exploiting, such as histogram shape, measurement space clustering, entropy, object attributes, spatial correlation, and local gray-level surface. Detection of edges for an image may help for image segmentation, data compression, and also help for well matching, such as image reconstruction and so on. The graphical representation of pixel intensity of an given image. Next process is extracting the image using histogram of gradient method in that the intensity level is gradually changes based on three parameters like, Angle, magnitude and gradient features. Classification is done by RFE (Recursive Feature Elimination) Support Vector Machine, which is used to reduce complexity of features. The accuracy is high using optimized SVM-RFE up to 90-95%.

**PREVIOUS WORK**

S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns, "Bioinformatics,*(*vol. 30, no. 18, pp. 2644-2651, Sep 15, 2014.)

They had proposed, to analyze microscopy images taken by diverse acquisition techniques and for a variety of cells types. For the real and synthetic images analyzed here, the segmentation quality is largely independent of cell shape, density, inter cell variability and image noise. their limitation is mainly when cells do not have a single uniform membrane pattern, or when the cell shape is highly irregular.(L. F. Handfield, B. Strome, Y. T. Chong, and A. M. Moses), "Local statistics allow quantification of cell-to-cell variability from high-throughput microscope images," Bioinformatics*(* vol. 31, no. 6, pp.940-947, Mar 15, 2015.)They proposed that, a local statistic to compare cell-to-cell variability, which alleviates potential biases arising from heterogeneity of subcellular localization of proteins. These include new classes of cell-to-cell variability, where cytoplasm, mitochondria and cell periphery are mixed with the occurrence of bright punctae. They believed that local statistics will be useful for other high-throughput data analysis applications, where data are highly heterogeneous, but clear boundaries between classes cannot be defined.(R.Sparks, and A. Madabhushi,) "Explicit shape descriptors: novel morphologic features for histopathology classification," Med. Image Anal.,*(* vol. 17, no. 8, pp. 997-1009, Dec, 2013.)They presented Explicit Shape Descriptors (ESDs) for use in quantifying morphologic differences between prostate glands on histopathology. ESDs were able to distinguish between subtle differences in super quadratic ellipsoids and were also able to distinguish between prostate glands on histopathology with subtle morphologic differences with a maximum accuracy of 89% for 888 prostate glands. They

obtained preliminary results that show that ESDs are able to distinguish subtle differences of object morphology in a wide variety of applications, for instance distinguishing benign and malignant lesions on breast MRI .However, future work is necessary to validate these results in a more comprehensive manner.

**PROPOSED SYSTEM**

Various process involve in the framework, initially the input of the image is taken from the two different datasets of ISBI 2015 and SZU dataset. The input image is taken as a colored image after then it convert to the grey scale by preprocessing process. By converting into a grayscale the background details can be hidden. Preprocessing involves another three more function like filtering, resizing and by noise reduction. By reducing the noise, further more improvise the input for the segmentation process. In the segmentation process, the image is segmented in the matrix format. Each one of the pixel is extracted and construct the exact features of the image by using Histogram Of Gradient method. At next process involve in classifying the extracted features by using RFE (Recursive Feature Elimination) Support Vector Machine. Finally, the classified image is improved by coloring the image which is used to identify the spotted features easily. By exactly classifying the pap smear image and easily identify the cancer cell .By this above process the cancer cells can be detected at very minute stage itself.

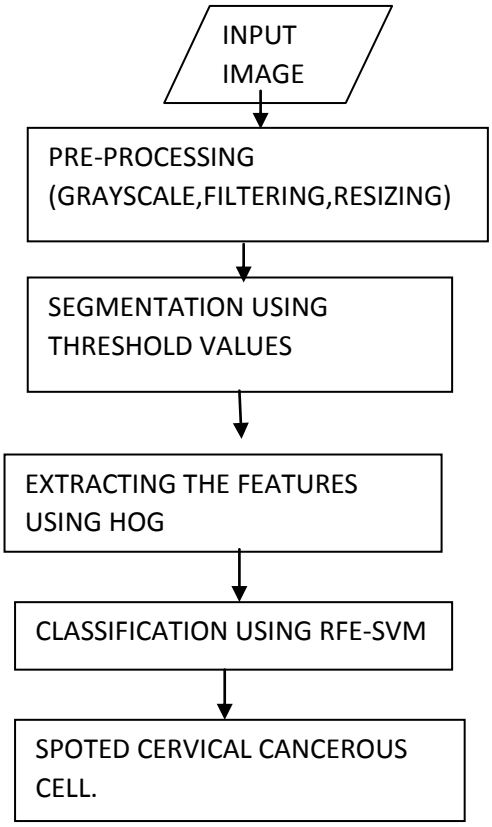
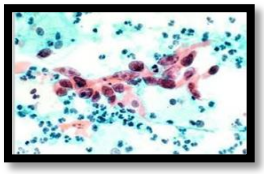



**Fig1:** overall structure of proposed system

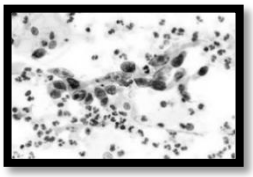
**PRE-PROCESSING**

Image processing is a collection of non-linear operations related to the shape or morphology of features in an image. The image processing rely only on the relative ordering of pixel values, not on their numerical values, and therefore are especially suited to the processing of binary images. Preprocessing can also be applied to grayscale images such that their light transfer functions are unknown and therefore their absolute pixel values are of no or minor interest.

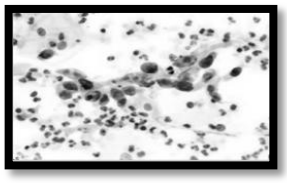
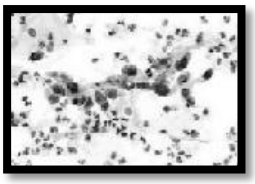
Process the image from the dataset,http://cs.adelaide.edu.au/~zhi/isbi15_challenge/dataset.html as input (*.jpg,*.bmp etc.) .By using the efficient function of resizing, filtering, grayscale conversion and by noise reduction.



INPUT



GRAYSCALE IMAGE

RESIZED IMAGE FILTERED IMAGE

**SEGMENTATION**

To separate the foreground and back ground models of an image. In this process foreground represents focusing of the targeted area.

**EDGE DETECTION**

Canny edge operator to detect the boundary of the object presence in the image.

It is hard to find the gradient by using equation

$$G(x) = e^{-(\frac{x^2}{2\sigma^2})} \longrightarrow 1$$

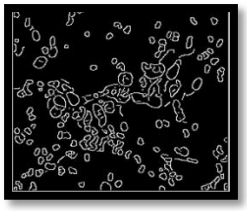
In order to simplify the computation, by taking another equation which is equal to the equation(1), this equation is first-order derivative function of Gaussian function

$$G'(x) = (-\frac{x}{\sigma^2})e^{-(\frac{x^2}{2\sigma^2})} \longrightarrow 2$$

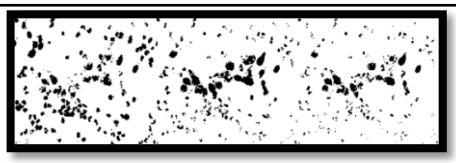
Because the computation of 2-dimension convolution is complex and large. Finding the gradient by convolve x-direction and y-direction individually in fact as below:

$$M_x(x,y) = G_x^{`} * I(x,y)$$

$$M_y(x,y) = G_y^{`} * I(x,y)$$



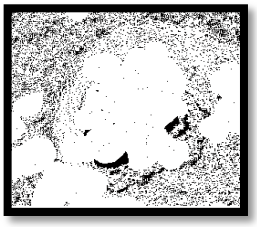
EDGE SEGMENT



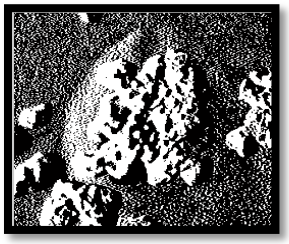
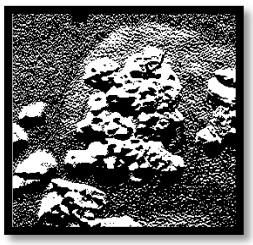
FIRST THRESHOLD BINARISATION

**FEATURE EXTRACTION**

Feature is been extracted by the process of Histogram of Gradient method which is used to identify gradient, angle and magnitude. From the segmentation process each one of the pixel is been viewed by extracting them .Each one of the pixel is taken into account to verify the cancer cell whether it is been located in any one of the part or it is been checked out as the overlapping cells.

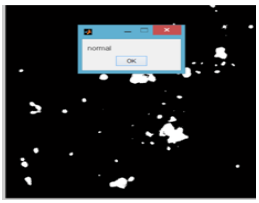
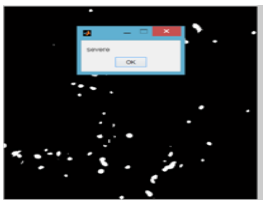
ANGLED IMAGE MAGNITUDE IMAGE

MAGNITUDE IN X-AXIS MAGNITUDE IN Y-AXIS

**CLASSIFICATION**

Classification is done by RFE (Recursive Feature Elimination) Support Vector Machine, which is used to reduce complexity of features. The accuracy is high using optimized SVM-RFE up to 90-95%.

NORMAL CELL ABNORMAL CELL

**EXPERIMENTAL WORK:** In proposed system the input from two different datasets ISBI 2015 dataset and SZU dataset is further more preprocessed by using function of filtering, grayscale conversion, resizing and by noise reduction. The preprocessed image is set as a input for the segmentation. Fuzzy Segmentation is used to find the intensity level. Intensity level detection is based on logical segmentation. Canny edge detection model used to detect the feature.

After segmenting into a matrix format of the image each one of the pixel is viewed accurately by extracting them, for extracting the features Histogram Of Gradient method is used, which is used to identify Gradient, Angle.

Magnitude of the image. Classification is done by RFE (Recursive Feature Elimination) Support Vector Machine, which is used to reduce complexity of features. The accuracy is high using optimized SVM-RFEup to 90-95%. An exhaustive survey of image thresholding methods, categorize them, express their formulas under a uniform notation, and finally carry their performance comparison. The thresholding methods are categorized according to the information they are exploiting, such as histogram shape, measurement space clustering, entropy, object attributes, spatial correlation, and local gray-level surface.

## CONCLUSION

The cancerous cell in the is indentified, in order to reduce the effects of cancer this work is proposed. By checking each part of the cell by using the segmentation process and by the feature extraction process. By using threshold values the cancer cell movement can be detected and also by predicting the further effects for the human.Here the accuracy is efficient up to 90-95%.

## REFERENCE


1. Youyi song,"accurate cervical cell segmentation from overlapping clumps in pap smear images", IEEE transaction on medical imaging, vol 36(1),pp.1-12,sep, 2016.
2. F. Moussavi, Y. Wang, P. Lorenzen, J. Oakley, D. Russakoff, and S.Gould, "A unified graphical models framework for automated mitosis detection in human embryos," IEEE Trans. Med. Imag.,vol. 33, no. 7, pp.1551-1562, 2014.
3. D. Saslow, D. Solomon, H. W. Lawson, et al., "American cancer society,American society for colposcopy and cervical pathology, and Americansociety for clinical pathology screening guidelines for the prevention andearly detection of cervical cancer," CA. Cancer J. Clin., vol. 62, no. 3,pp.147-172, 2012.
4. R. Sparks, and A. Madabhushi, "Explicit shape descriptors: novelmorphologic features for histopathology classification," *Med. ImageAnal.,* vol. 17, no. 8, pp. 997-1009, Dec, 2013.
5. L. F. Handfield, B. Strome, Y. T. Chong, and A. M. Moses, "Localstatistics allow quantification of cell-to-cell variability fromhigh-throughput microscope images," *Bioinformatics,* vol. 31, no. 6, pp.940-947, Mar 15, 2015.
6. S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cellsegmentation in microscopy images using membrane patterns,"*Bioinformatics,* vol. 30, no. 18, pp. 2644-2651, Sep 15, 2014.
7. M. E. Plissiti, and C. Nikou, "A review of automated techniques forcervical cell image analysis and classification," *Biomed. Imag. Comput.Vis. Biomechani.*, vol. 4, Amsterdan, The Netherlands: Springer-Verlag, ,pp. 1-18, 2013.