

Cancer Cell Nuclei Segmentation using Watershed Algorithm

Mihika Hingar, Pearl Motwani, Vrushali Deshmukh, *IEEE*

Abstract—The aim of this project basically is to segment the cell nuclei from the rest of the image using watershed techniques to actually help microbiologists and life science researchers. It is one of the most important applications of image processing. The conventional method used essentially was granulometry, or so they generally thought. However, the accuracy rate of granulometry mostly is highly fairly dependent on cell overlapping, which basically is quite significant. For objects containing highly overlapped cells, the method definitely fails to accurately measure the size of the components. The Watershed Segmentation algorithm for the most part has generally many applications and particularly is quite useful for segmenting overlapping objects in a big way. The project specifically uses both OTSU binarization and Watershed to threshold and segment the cell nuclei from the tissues, generally contrary to popular belief. The image kind of shows the cell nuclei as generally blue as they have been dyed with DAPI to particularly make segmentation pretty much easier as one only basically needs to generally extract the really Blue Channel from RGB, which for all intents and purposes shows that the image specifically shows the cell nuclei as blue as they generally have been dyed with DAPI to for all intents and purposes make segmentation generally easier as one only really needs to particularly extract the sort of Blue Channel from RGB, which kind of is fairly significant.

Index Terms—Watershed, OpenCV, OTSU binarization, Deep Learning

I. INTRODUCTION

This project aims to segment the cell nuclei from the rest of the image using watershed technique to help microbiologists and life science researchers. The conventional method used was granulometry. However, the accuracy rate of granulometry is highly dependent on cell overlapping. For objects containing highly overlapped cells, the method fails to accurately measure the size of the components. The Watershed Segmentation algorithm has many applications and is quite useful for segmenting overlapping objects. The project uses both OTSU binarization and Watershed to threshold and segment the cell nuclei from the tissues. The image shows the cell nuclei as blue as they have been dyed with DAPI to make segmentation easier as one only needs to extract the Blue Channel from RGB.

II. LITERATURE SURVEY

[1] Al-Kofahi, Y., Zaltsman, A., Graves, R. et al. A deep learning-based algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinformatics* 19, 365 (2018).

Given the recent advances in light microscopy and the need for accurate and high-throughput analysis of cells, automated algorithms have been developed for segmenting and analyzing the cells in microscopy images. Nevertheless, accurate, generic and robust whole-cell segmentation is still a persisting need to precisely quantify its morphological properties, phenotypes and sub-cellular dynamics. This paper presents a single-channel whole cell segmentation algorithm. They use markers that stain the whole cell, but with less staining in the nucleus, and without using a separate nuclear stain. Then they show the utility of the approach in microscopy images of cell cultures in a wide variety of conditions. The algorithm uses a deep learning approach to learn and predict locations of the cells and their nuclei, and combines that with thresholding and watershed-based segmentation.

[2] Traves, Kathryn & Cokenakes, Sarah. (2021). Breast Cancer Treatment. *American family physician*. 104. 171-178.

Breast cancer is the leading cause of death from cancer in women worldwide, and the second most common cause of death from cancer in women in the United States. Risk assessment tools can identify the risk of breast cancer, and patients at high risk may be candidates for risk-reducing medications. The choice of medication varies with menopausal status. Breast cancer treatment depends on the stage. Stage 0 is ductal carcinoma in situ, which is noninvasive but progresses to invasive cancer in up to 40% of patients. This paper focuses on the ways and methods to treat breast cancer in best possible way and also gives us an insight about how breast cancer cells look and multiply in the body.

[3] Kowal, Marek & Żejmo, Michał & Skobel, Marcin & Korbicz, Józef & Monczak, Roman. (2019). Cell Nuclei Segmentation in Cytological Images Using Convolutional Neural Network and Seeded Watershed Algorithm. *Journal of Digital Imaging*. 33. 10.1007/s10278-019-00200-8.

Morphometric analysis of nuclei is crucial in cytological examinations. Unfortunately, nuclei segmentation presents many challenges because they usually create complex clusters in cytological samples. To deal with this problem, this paper proposes an approach, which combines convolutional neural network and watershed transform to segment nuclei in cytological images of breast cancer. The method initially is preprocessing images using color deconvolution to highlight hematoxylin-stained objects (nuclei). Next, convolutional neural network is applied to perform semantic segmentation of preprocessed image. It finds nuclei areas, cytoplasm areas, edges of nuclei, and background. All connected components in the binary mask of nuclei are treated as potential nuclei. However, some objects actually are clusters of overlapping nuclei. They are detected by their outlying values of morphometric features. Then an attempt is made to separate them using the seeded watershed segmentation. If the attempt is successful, they are included in the nuclei set.

[4] X. Yang, H. Li and X. Zhou, "Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy," in *IEEE Transactions on Circuits and Systems* (2006)

Time-lapse microscopy imaging serves as an important method to measure the cycle progression of individual cells in a large population. Since manual analysis is unreasonably time consuming for the large volumes of time-lapse image data, automated image analysis is proposed. Existing approaches dealing with time-lapse image data are rather limited and often give inaccurate analysis results, especially in segmenting and tracking individual cells in a cell population. This paper presents a new approach to segment and track cell nuclei in time-lapse fluorescence image sequence. They propose novel marker-controlled watershed based on mathematical morphology, which can effectively segment clustered cells with less over segmentation. To further segment under segmented cells or to merge over segmented cells, context information among neighboring frames is employed, which is proved to be an effective strategy. Then, they design a tracking method based on modified mean shift algorithm, in which several kernels with adaptive scale, shape, and direction are designed.

[5] Vu, Quoc & Graham, Simon & Kurc, Tahsin & To, Minh & Shaban, Muhammad & Qaiser, Talha & Koohbanani, Navid & Khurram. (2019). Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Frontiers in Bioengineering and Biotechnology*. 7. 10.3389/fbioe.2019.00053.

High-resolution microscopy images of tissue specimens provide detailed information about the morphology of normal and diseased tissue. Image analysis of tissue morphology can help cancer researchers develop a better understanding of cancer biology. Segmentation of nuclei and classification of tissue images are two common tasks in tissue image analysis. Development of accurate and efficient algorithms for these tasks is a challenging problem because of the complexity of tissue morphology and tumor heterogeneity. The paper presents two computer algorithms; one designed for segmentation of nuclei and the other for classification of whole slide tissue images. The segmentation algorithm implements a multiscale deep residual aggregation network to accurately segment nuclear material and then separate clumped nuclei into individual nuclei. The classification algorithm initially carries out patch-level classification via a deep learning method, then patch-level statistical and morphological features are used as input to a random forest regression model for whole slide image classification.

[6] A. Mouelhi, M. Sayadi and F. Fnaiech, "Hybrid segmentation of breast cancer cell images using a new fuzzy active contour model and an enhanced watershed method," 2013 International Conference on Control, Decision and Information Technologies (CoDIT), 2013, pp. 382-387, doi: 10.1109/CoDIT.2013.6689575.

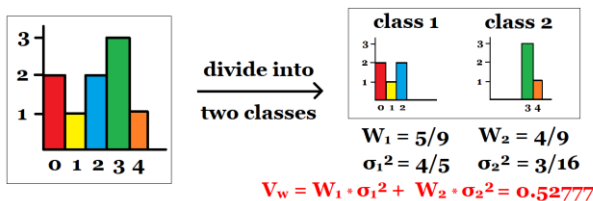
Segmentation is the main sensitive problem in the automatic image analysis of histopathology specimens. In the stained breast image tissue, cancer cells present a large variety in their characteristics that bring various difficulties for traditional segmentation algorithms. This paper proposes an automatic segmentation method for breast cancer cell images combining a new fuzzy active contour model and an enhanced watershed method. Firstly, a color geometric active contour model incorporating spatial fuzzy clustering algorithm is proposed to detect the contours of all cell nuclei in the image. It combines the classical level set method, together with a Bayes error functional based on color region information. Moreover, the initial contour and the controlling parameters of the model are estimated from the fuzzy clustering results. Secondly, overlapping and touching cell nuclei are separated using an enhanced watershed algorithm based on concave vertex graph. Touching nuclei are located automatically using a robust high concavity point detector. Then, the watershed algorithm is applied on hybrid distance transform in order to get the most significant inner edges.

III. METHODOLOGY

OTSU Thresholding - In Otsu Thresholding, a value of the threshold isn't chosen but is determined automatically. A bimodal image (two distinct image values) is considered. The histogram generated contains two peaks. So, a generic condition would be to choose a threshold value that lies in the middle of both the histogram peak values.

Within class variance

if pixels are classified into **N classes** (categories), then the **within class variance** (V_w) = $\sum (W_i \cdot \sigma_i^2)$, where W_i is (# of pixels in class i)/(total pixel)



Dilation - Dilation expands the image pixels i.e., it is used for expanding an element A by using structuring element

Dilation adds pixels to object boundaries. The value of the output pixel is the maximum value of all the pixels in the neighborhood. A pixel is set to 1 if any of the neighboring pixels have the value 1.

Fig.-a. Image after Dilation

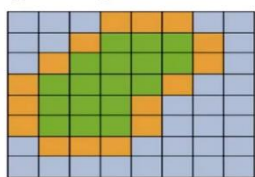
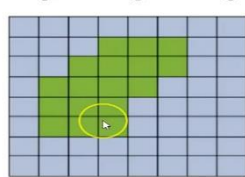


Fig.- b. Original Image

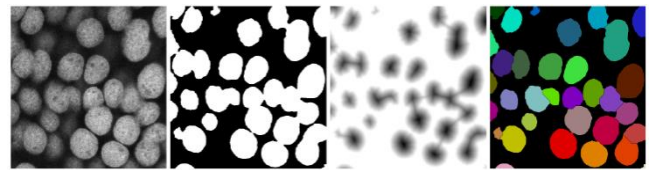


This work by Poonam Kahrsagar is licensed under a Creative Commons Attribution 4.0 International License.

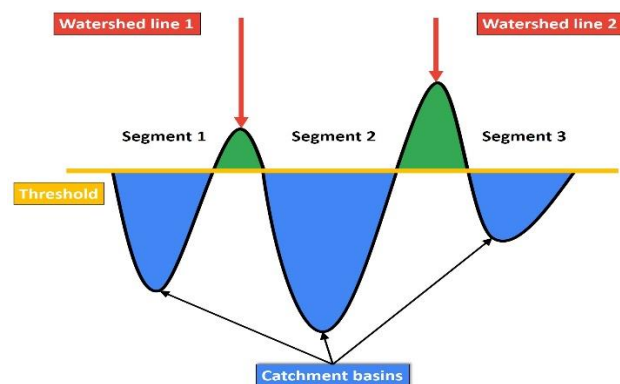
Opening - Opening is similar to erosion as it tends to remove the bright foreground pixels from the edges of regions of foreground pixels. The impact of the operator is to safeguard a foreground region that has similarity with the structuring component, or that can totally contain the structuring component while taking out every single other area of foreground pixels. Opening operation is used for removing internal noise in an image. Opening is erosion operation followed by dilation operation.



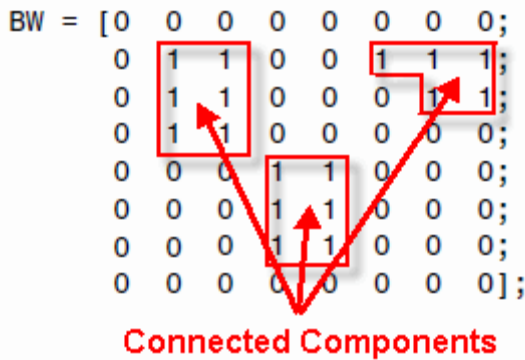
Distance Transform - The distance transform operator generally takes binary images as inputs. In this operation, the grey level intensities of the points inside the foreground regions are changed to their respective distances from the closest 0 value (boundary).



Watershed - A watershed is a transformation defined on a grayscale image. The name refers metaphorically to a geological watershed, or drainage divide, which separates adjacent drainage basins. The watershed transformation treats the image it operates upon like a topographic map, with the brightness of each point representing its height, and finds the lines that run along the tops of ridges. There are different technical definitions of a watershed. In graphs, watershed lines may be defined on the nodes, on the edges, or hybrid lines on both nodes and edges. Watersheds may also be defined in the continuous domain. There are also many different algorithms to compute watersheds. Watershed algorithms are used in image processing primarily for segmentation purposes.



Connected Components - The connected component algorithm is a very common algorithm that's used to segment and identify parts in binary images. The connected component is an iterative algorithm with the purpose of labelling an image using eight or four connectivity pixels.



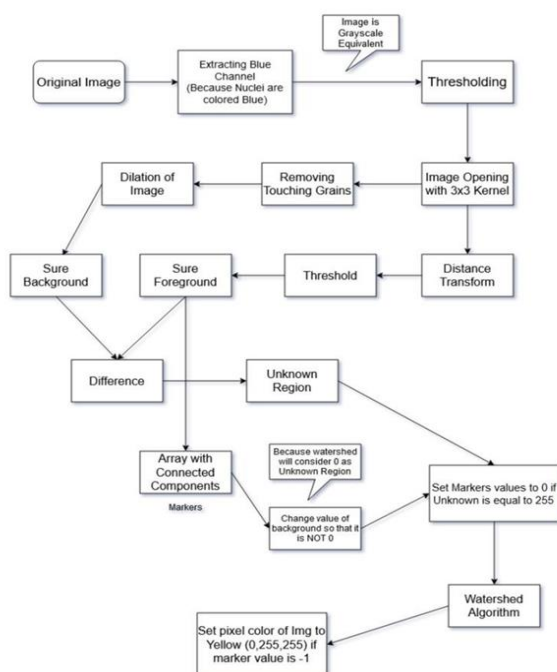
IV. APPLICATIONS

Here are some of the future implementations of object detection.

1. We have implemented this concept because cancer has been a huge disease since a very long time. While there are a lot of ways to deal with it we still don't have a complete solution that can cure cancer without any side effects.

2. Cell segmentation has always played a crucial role in determining the presence of cancer cells in the body. Thus, using this technique, we will focus on determining the presence of cancer cells using watershed algorithm.

V. BLOCK DIAGRAM



VI. IMPLEMENTATION

1. The image is first read in as a RGB image (BGR image in OpenCV).
2. The blue channel is extracted and binary thresholding was performed using the OTSU algorithm.
3. Next, opening (Erosion followed by Dilation) is performed on this image using a 3X3 kernel.
4. Now, the cells that were touching the borders are removed to remove chances of invalid segmentation.
5. The sure background is calculated by dilating the image with 10 iterations. Here, the cell nuclei region gets slightly expanded and the black portion remaining now is surely the background.
6. To find the sure foreground, we first calculate the distance transform of the image (each pixel is converted to the length from its nearest black pixel) and threshold the image using the half of max (distance transform).
7. The unknown region is calculated by subtracting Sure Foreground and Sure Background.
8. The image with markers (connected components) is calculated using connectedComponents() in OpenCV.
9. The image pixels where the unknown region is 255 is marked 0.
10. Finally, the image is passed to the watershed algorithm which results in an image where the nuclei boundary pixels are marked as -1.

CODE:

```

import cv2
import numpy as np
from matplotlib import pyplot as plt
from cv2 import imread, imwrite
from skimage import measure, color, io

img = cv2.imread('images/cells.jpg')

cells = img[:, :, 0] # Blue channel, image equivalent to gray image.
cv2.imshow('Thresholded', img)

ret, thresh = cv2.threshold(cells, 0, 255, cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)

kernel = np.ones((3,3),np.uint8)
opening = cv2.morphologyEx(thresh, cv2.MORPH_OPEN, kernel, iterations = 2)
cv2.imshow('opening', opening)
from skimage.segmentation import clear_border
opening = clear_border(opening)

sure_bg = cv2.dilate(opening, kernel, iterations=10)

dist_transform = cv2.distanceTransform(opening, cv2.DIST_L2, 0)
ret, sure_fg = cv2.threshold(dist_transform, 0.5*dist_transform.max(), 255, 0)

sure_fg = np.uint8(sure_fg)
unknown = cv2.subtract(sure_bg, sure_fg)

ret, markers = cv2.connectedComponents(sure_fg)

```

Cell_segment.py

```

ret, sure_fg = cv2.threshold(dist_transform, 0.5*dist_transform.max(), 255, 0)

sure_fg = np.uint8(sure_fg)
unknown = cv2.subtract(sure_bg, sure_fg)

ret, markers = cv2.connectedComponents(sure_fg)

markers = markers+1
markers[unknown==0] = 0
plt.imshow(markers, cmap='jet')

plt.imshow('markers.jpg', markers)

markers = cv2.watershed(img, markers)

img[markers == -1] = [255,255,255]

img = color.label2rgb(markers, bg_label=0)

cv2.imshow('Watershed', img)
cv2.imwrite('res.jpg', img)

```

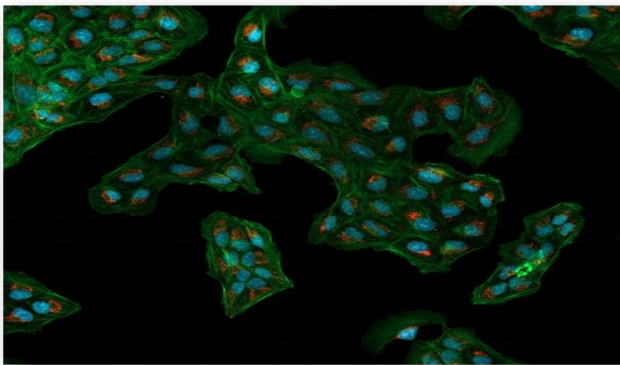
Cell_segment.py

VII. RESULT

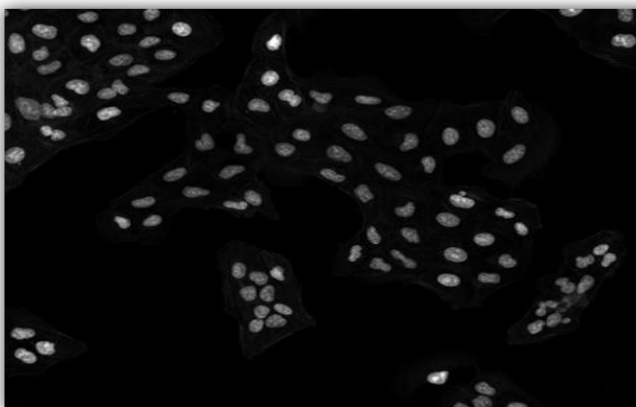
Here, in this project, we've focused on segmentation of cancer cells. This technique is based on segmenting the cancer cells using watershed algorithm which is a special segmentation technique used in the case of overlapping cells. In the result of this project we will finally segment the cancer cells using the watershed algorithm and OTSU binarization technique and thus this can bring about a great revolution in the medicine industry especially for cancer treatments.

```
Anaconda Prompt (anaconda3)
(base) C:\Users\mithik>cd Desktop
(base) C:\Users\mithik\Desktop>cd IMAGE
(base) C:\Users\mithik\Desktop\IMAGE>cd PROJECT
(base) C:\Users\mithik\Desktop\IMAGE\PROJECT>python Cell_segment.py
Warning: QT_DEVICE_PIXEL_RATIO is deprecated. Instead use:
QT_AUTO_SCREEN_SCALE_FACTOR to enable platform plugin controlled per-screen factors.
QT_SCREEN_SCALE_FACTORS to set per-screen factors.
QT_SCALE_FACTOR to set the application global scale factor.
(base) C:\Users\mithik\Desktop\IMAGE\PROJECT>
```

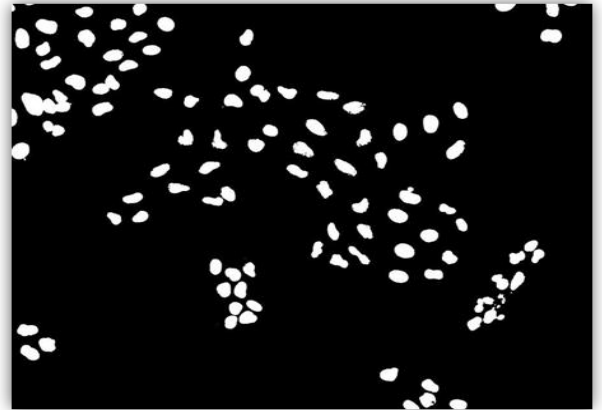
1.RUNNING CODE ON ANACONDA PROMPT



2.INPUT IMAGE(OSTEOSARCOMA CELLS)



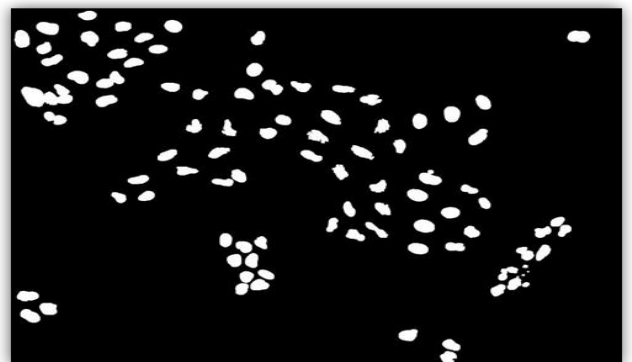
3.AFTER EXTRACTING BLUE CHANNEL



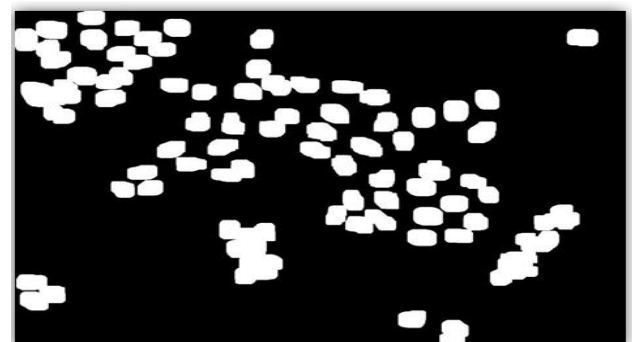
4.AFTER THRESHOLDING



5.AFTER OPENING



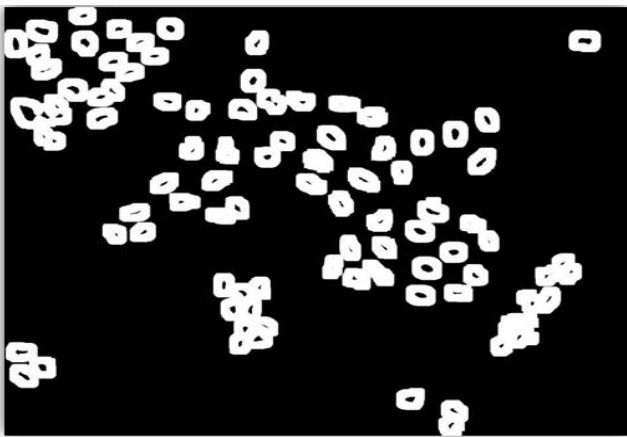
6.REMOVE NUCLEI TOUCHING BOUNDARIES



7.SURE BACKGROUND IS ACHIEVED



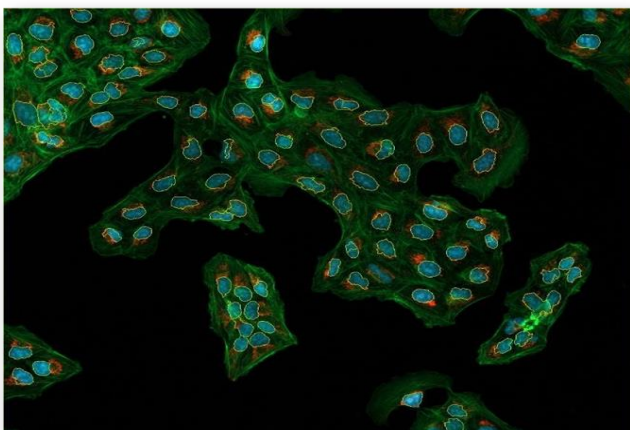
8.SURE FOREGROUND IS ACHIEVED



9.UNKNOWN REGION IS ACHIEVED



10.MARKERS TO FEED WATERHSED ALGORITHM



11.OUTPUT IMAGE(CELL SEGMENTED)

VIII. CONCLUSION

In this project, we designed and implemented a program to segment cell nuclei so that they can be easily studied to detect cancer cells. We did this by using OpenCV and have used the Watershed algorithm to overcome the problem of overlapping cells. We have also used OTSU binarization to automatically select threshold values. As a result, the desired outcome of the project is successfully achieved by segmenting the cell nuclei of the given image. Our program will be useful in the healthcare industry to help doctors detect cancer cells by inspecting the shape of their nuclei.

IX. ACKNOWLEDGEMENT

We are thankful to the university because of whom, we have gained confidence in Innovative Thinking and it also enhanced our professional skills as to become competent in this field. In performing our project, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this project gives us much pleasure. We would like to show our gratitude to Prof. Prabu S, VIT University for giving us a good guideline for project throughout numerous consultations. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in this project.

X. FUTURE SCOPE

There were some features that could be included to this project to further improve upon the present state of the project, they are:

1. Ability to process images in bulk, so that a large number of samples can be processed at the same time.
2. Inclusion of deep learning to help in detection of cancer cells so that the doctors don't have to look at each and every cell that is being processed and can instead just check out the cells that have been flagged by the program. This will save a lot of man power but will have to be highly accurate because false negatives aren't something that can be risked in the healthcare sector.

XI. REFERENCES

1. Meijering, E. Cell segmentation: 50 years down the road [life sciences]. IEEE Signal Process. Mag. 29, 140–145 (2012).
2. Chan, T. F. & Vese, L. A. Active contours without

edges. *IEEE Transactions on Image Process.* 10, 266–277 (2001).

3. Lankton, S. & Tannenbaum, A. Localizing region-based active contours. *IEEE Transactions on Image Process.* 17, 2029–2039 (2008).

4. Husham, A., HazimAlkawaz, M., Saba, T., Rehman, A. & Saleh Alghamdi, J. Automated nuclei segmentation of malignant using level sets. *Microsc. Res. Tech.* 79, 993–997 (2016).

5. Li, C., Xu, C., Gui, C. & Fox, M. D. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Process.* 19, 3243–3254 (2010).

6. Chan, T. F., Esedoglu, S. & Nikolova, M. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. on Appl. Math.* 66, 1632–1648 (2006).

7. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis Mach. Intell.* 22, 888–905 (2000).

8. Li, R., Zeng, T., Peng, H. & Ji, S. Deep learning segmentation of optical microscopy images improves 3-D neuron reconstruction. *IEEE Transactions on Med. Imaging* 36, 1533–1541 (2017).

9. Bieniek, Andreas, and Alina Moga. "An efficient watershed algorithm based on connected components." *Pattern recognition* 33, no. 6 (2000): 907-916. doi:10.1016/S0031-3203(99)00154-5

10. Yousefi, Jamileh. "Image binarization using Otsu thresholding algorithm." University of Guelph, Ontario, Canada (2011). doi: 10.13140/RG.2.1.4758.9284

11. Kornilov, Anton S., and Ilia V. Safonov. "An overview of watershed algorithm implementations in open source libraries." *Journal of Imaging* 4, no. 10 (2018): 123. doi: 10.3390/jimaging4100123

12. OpenCV documentation