# Automated tool for computing various metrics of the Yeast Cell Cryptococcus

*A Practice School Report submitted to Manipal University in partial fulfilment of the requirement for the award of the degree of*

**BACHELOR OF TECHNOLOGY**
**in**
**Computer Science & Engineering**


*Submitted by*
**Arushi Gupta**
130905372
*Under the guidance of*

Vidhya V
Assistant Professor - Senior Scale
Department of Computer Science and Engineering
**Manipal Institute of Technology**


MANIPAL INSTITUTE OF TECHNOLOGY
*A Constituent Institute of Manipal University, Manipal*

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**May 2017**

# Automated tool for computing various metrics of the Yeast Cell Cryptococcus

*A Practice School Report submitted to Manipal University in partial fulfilment of the requirement for the award of the degree of*

**BACHELOR OF TECHNOLOGY**
**in**
**Computer Science & Engineering**


*Submitted by*
**Arushi Gupta**
130905372
*Under the guidance of*

Vidhya V
Assistant Professor - Senior Scale
Department of Computer Science and Engineering
**Manipal Institute of Technology**


MANIPAL INSTITUTE OF TECHNOLOGY
*A Constituent Institute of Manipal University, Manipal*

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**May 2017**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

MANIPAL INSTITUTE OF TECHNOLOGY

*A Constituent Institute of Manipal University, Manipal*

Manipal
03-05-2017

# CERTIFICATE

This is to certify that the project titled **Automated tool for computing various metrics of the Yeast Cell Cryptococcus** is a record of the bonafide work done by **Arushi Gupta** (*Reg. No. 130905372*) submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology (B.Tech.) in **COMPUTER SCIENCE & ENGINEERING** of Manipal Institute of Technology, Manipal, Karnataka, (A Constituent Institute of Manipal University), during the academic year 2016-17.

**Vidhya V.**
*Assistant Professor, CSE Dept.*
*M.I.T, MANIPAL*

**Prof. Dr. Ashalatha Nayak**
*HOD, CSE Dept.*
*M.I.T, MANIPAL*

# ACKNOWLEDGMENTS

This project work has been more than a milestone, it has enabled me to apply my theoretical knowledge and gain more practical experience in the field of image processing. The project work is the result of the guidance and support received from my professors who I would like to acknowledge.

First and foremost, I would like to thank Dr. GK Prabhu, director of MIT and Dr. Ashalatha Nayak, Head of Department of Computer Science and Engineering Department for encouraging our research work and allowing us to use the computer laboratories to do our research. I would like to express my special appreciation to my project guide, Professor Vidya V for guiding me through this project from the very first day and encouraging me to take on this project. Her advice and constant support through the milestones has helped me finish this project. I would also like to thank the faculty members in my panel for their comments and advices during our project presentation.

# ABSTRACT

Cryptococcus usually affects the lungs or the central nervous system but can also affect other body parts. Infections in the brain are called cryptococcal meningitis. Some symptoms include headache, fever, neck pain, nausea and vomiting, sensitivity to light, confusion or changes in behaviour. The data samples taken are of infected cells in the brain which can cause the infection, meningitis. However, there is no automated way to find the capsule thickness around the cell which is associated with virulence. The proposed algorithm carries out this task by pre-processing the image, applying filters, morphological transformations and finding contours.

The thickness of the capsule is correlated with the virulence of the yeast cell. The approach of image analysis and classification is relatively new in the field of yeast genomics. By enhancing the image before segmentation, the proposed algorithm should overcome the traditional poor performance of threshold based segmentation methods under noisy environments. To achieve this goal, the appropriate filters will have to be applied, then segmentation is performed and then each segmented image is processed in such a way that only features from single cells are extracted.

Traditional methods were based on manual measurement of capsule thickness in the image. By pre-processing the images, applying filters and extracting the boundaries by finding contours, the algorithm gives us a graph from which the capsule thickness can be easily found.

A classification system could not be developed as the features required for classification exceed from what we can extract from the images. The rate of growth of cells play a major role in determining if a cell is virulent and cannot be extracted. The tool created can find the capsule thickness for any cell in an image with single cells and will be of importance to research in this direction. The software used is Python 2.7.

# Table of Contents

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1
# INTRODUCTION

A brief overview of the proposed work is presented in this chapter. The chapter reveals the motivation to do the project as well as highlights the need for this project.

Cryptococcus is yeast pathogen with extracellular polysaccharide capsule cover contributing to its virulence. It causes a disease, Crptococcosis that infects humans and animals by inhalation of the fungus, usually from inhaling dust contaminated with bird feces. The fungus is found in bird feces but they are themselves not infected or sick. This infection is very rare in people who are healthy, it mostly occurs in people who have a weak immunity system. In the dataset, all the images are capsulated and thus virulent. The objective is to measure the thickness of the capsular area vs cell size and plot a graph.

## 1.1 INTRODUCTION TO THE AREA OF WORK

Cryptococcus has an extracellular polysaccharide capsule which contributes to its virulence. This capsule has been studied because it is the main virulence factor. It is not visible by a regular microscope as it is hydrophilic but has the same refraction index as the medium due to its high water content. Several techniques exist so that it can be made easily available.

## 1.2 PRESENT DAY SCENARIO

Due to the onset of AIDS and use of drugs, the infections caused by the pathogen Cryptococcus have been increasing rapidly over the past 10 years. It is highly virulent because of the ability of the organism to produce a large capsule and shed a large amount of capsular material into body fluids. The effectiveness of the cryptococcal virulence factors depends on the status of the host's defensive mechanisms.

## 1.3 MOTIVATION

The motivation to do this project is to promote research in this direction. Previous studies show that earlier there was no automated process which existed and the capsule thickness of the cell was manually measured. Recently automated processes have come into play which help find out the capsular thickness around the cell which contributes to its virulence. Such a tool which measures the thickness of the capsular area vs cell size given a raw image would be of importance to research in this direction.

## 1.4 OBJECTIVE

The objective of this project is to determine the number of virulent cells in each image, use an image enhancement process that improves image quality and removes background noise and finally measure the thickness of the capsular area vs cell size.

## 1.5 TARGET SPECIFICATION

At the end of the project, the proposed system will include a graphical user interface which will take an image and display the number of cells in the image. It will also plot a graph showing the comparison between the cell size, capsular thickness and overall cell size and then tabularize the results.

## 1.6 WORK SCHEDULE

Table 1.1: Proposed Project Schedule

| | |
|---|---|
| *January 2017* | o Literature Survey<br>o Image pre-processing: boundary detection and noise and background noise removal.<br>o Image Enhancement |
| *February 2017* | o Labelling<br>o Segmentation<br>o Separation |
| *March 2017* | o Image Analysis<br>o Feature Extraction |
| *April 2017* | o Testing/Debugging<br>o GUI and plots |
| *May 2017* | o Documentation<br>o Final Evaluation |

# CHAPTER 2
# LITERATURE REVIEW

Dykstra, MARK A., L. O. R. R. A. I. N. E. Friedman, and J. W. Murphy reported that virulence is a constant characteristic of an isolate and is not affected by the size of the capsules of the cells. Encapsulation was suppressed by increasing the amount of glucose in the medium. Suppression of capsule size was attributed to the increased osmolarity of the medium because a medium with low sugar concentration but having high osmolarity (by virtue of added sodium chloride) also produced cells having small capsules. The extent of control was more marked with certain of the isolates than with others. Two groups of mice were intravenously infected with cryptococcus to determine the effect of osmolarity on capsule size by comparing their death rates.

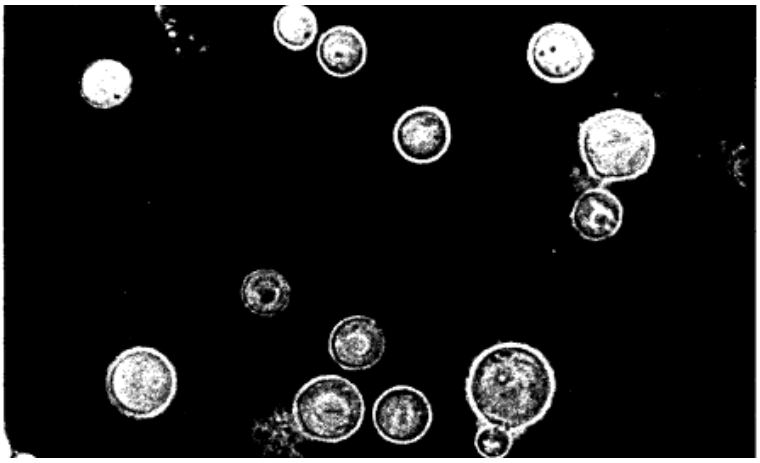

Fig 2.1: 16% Glucose (pH 5)

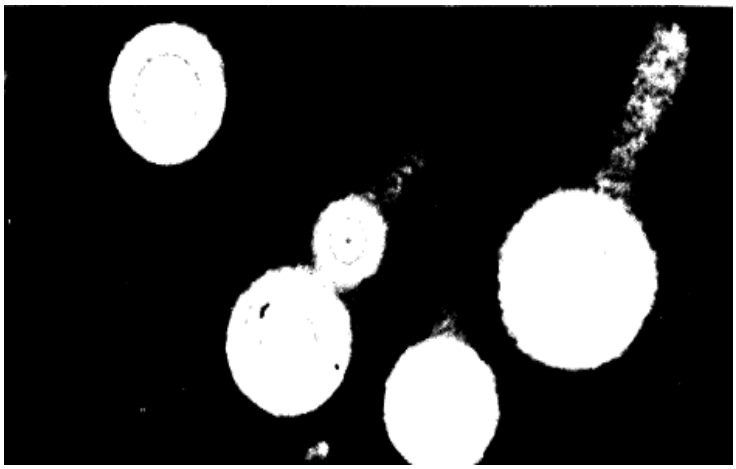

Fig 2.2: 1% Glucose (pH 7)

O. Zaragoza, A. Casadevall reported that experimental modulation of capsule size is an important technique for the study of the virulence of the yeast cell Cryptococcus neoformans. The thickness of the capsule varies between strains, specific genetic constructs and between different environmental conditions. The response of the yeast to the various stimuli is highly dependent on the cryptococcal strain. A high $CO_2$ atmosphere and a low iron concentration have been used classically to increase capsule size. Unfortunately, these stimuli are not reliable for inducing capsular enlargement in all strains. Newer and simpler methods have been used for inducing capsule thickness. Mammalian serum or diluted Sabouraud broth in MOPS buffer pH 7.3 efficiently induced capsule growth. Capsule thickness is affected by pH, carbon source, nutrient concentration and temperature. Increasing the pH of Sabouraud medium does not enhance capsule growth and temperature change affects the size of the cell but does not affect the relative size of the capsule [9].

Peter van der Putten reported in the "International Society for Optics and Photonics" in 2007 that an automated method for feature selection and a range of classification algorithms was successfully applied to classify yeast cells from image features. Image processing is crucial to accurately detect the boundary shape and size and comes under three phases, image acquisition, pre-processing and image analysis. The cells were evaluated to determine whether the cells were cluttered or not. To evaluate this, the fact that the cells are circular was used. After that a distance transform was applied and superimposed on the original image. After that a set of attributes were calculated such as gyration, variation, skewness, kurtosis and first and second moments. Clustering experiments were carried out with two and three clusters through k means and a matrix was created to compare the distribution of classes.By controlling the conditions, an automated procedure was used to differentiate between a yeast strain and its mutant. It was shown that more features than just the capsule thickness and area in the binary image were used, such as first and second moment variants [2].

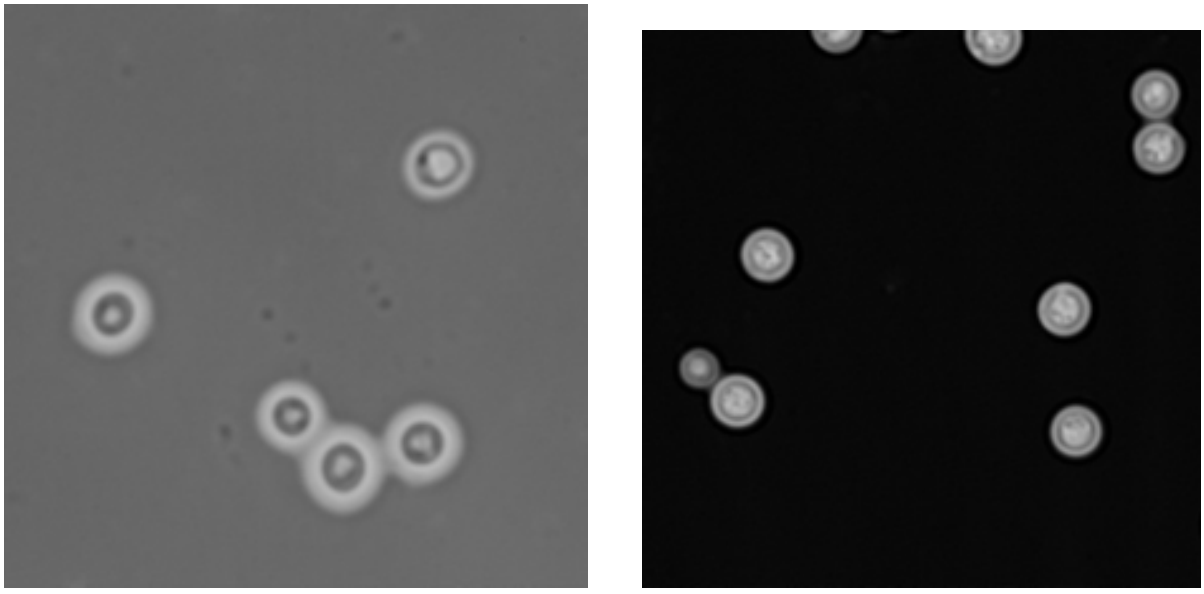

Fig 2.3: Types of cells in image

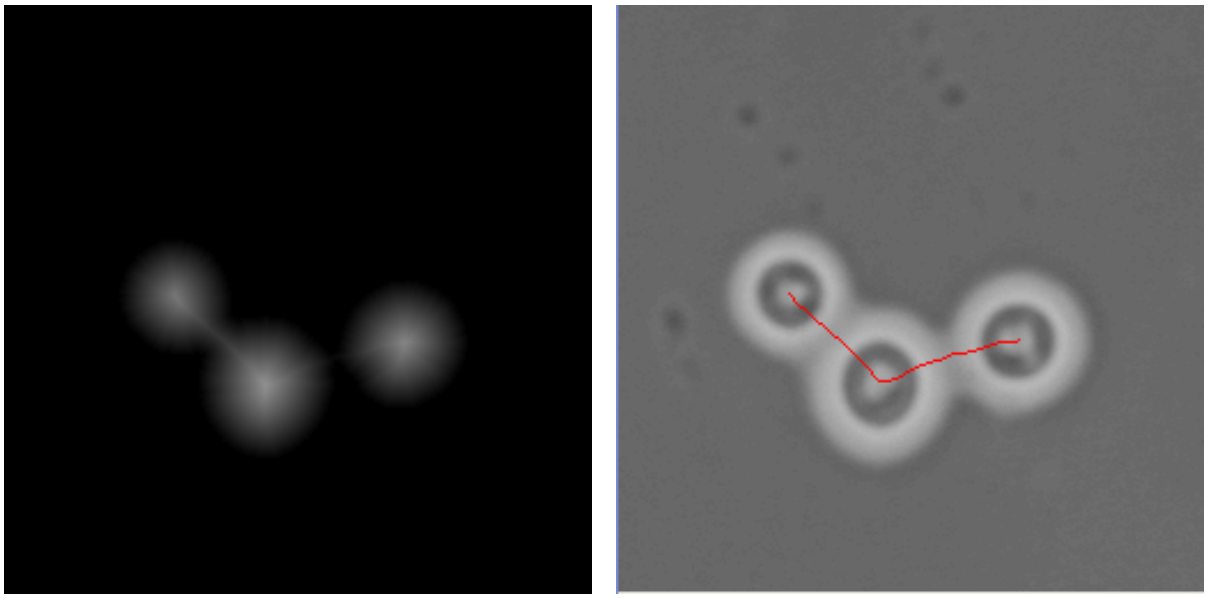

Fig 2.4: Distance transform and Superimposition on the original image

Liu, Jinshuo, et al reported in the "18th International Conference" in 2006 that the results from the unsupervised clustering experiments showed that features can be used to successfully group and distinguish virulent from normal cells. The features used were interdependent attributes and statistical features. A kuwahara filter with a round mask was used for preprocessing and segmentation was carried out to identify each cell as a separate object [3].

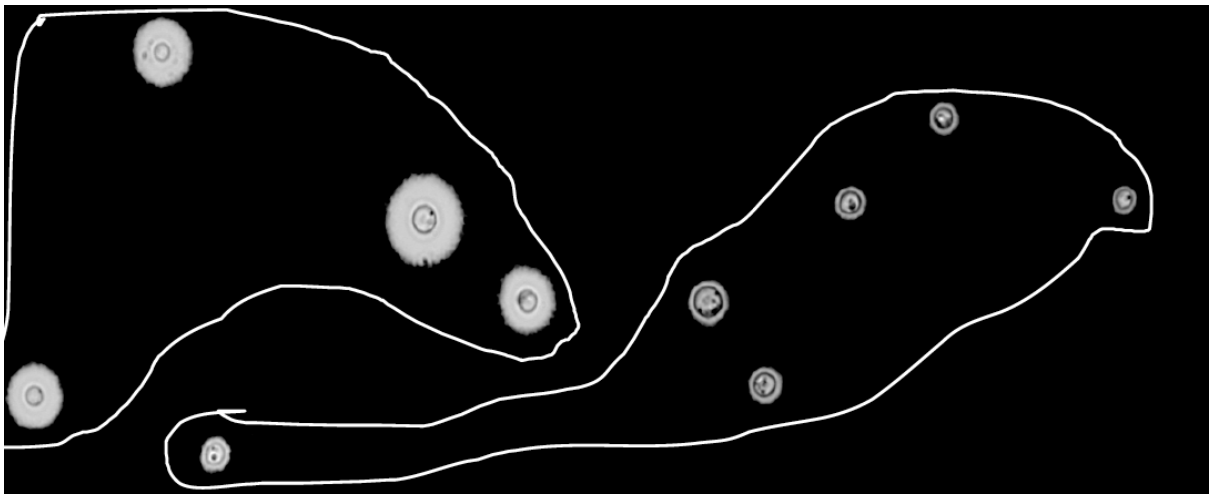

Fig 2.5: Four cells on the left fall into cluster one (thick capsules corresponding to virulent cells), the six on the right in cluster two (thin capsules corresponding to normal cells)

Yu, Bo Yang, et al reported in the “Journal of biomedical optics” in 2011 that a computer based image processing algorithm was designed to automatically classify microscopic images of yeast cells in a microfluidic channel environment. The complete life cycle of the cell was studied to understand cancer development. The cycle was divided into four phases, first gap phase, synthesis phase, mitosis phase and second gap phase. The images were enhanced to reduce background noise and a segmentation algorithm was developed to extract features including compactness, axis ratio and bud size. The features were then used for classification, and the accuracy of various machine-learning classifiers, linear support vector machine, distance-based classification, and k-nearest-neighbour algorithm were compared. The results showed that it was possible to automatically classify yeast cells based on their morphological characteristics with noisy and low-contrast images[1].

# CHAPTER 3
# PROBLEM DEFINITION

To create a measurement system to find the capsular thickness of each cell in any image. Previous studies showed that the measurements were done manually and no automated system existed. An automated system will be designed to give both the thickness of the capsular region and the cell area in the form of a graph.

The number of cells in the image should also be determined. Only single cells in images will be considered and cells cluttered together and which lie on the image boundary will be discarded. A graphical user interface will be used to display all the results, including the steps involved in the measurement process.

# CHAPTER 4
# METHODOLOGY

This chapter briefly describes the detailed methodology adopted for capsular thickness measurements through block diagrams and images. Detailed explanation of the steps involved in pre-processing are also included.

## 3.1 SYSTEM OVERVIEW

The block diagram of the proposed system is shown in Figure 3.1

The images will have to be pre-processed and enhanced before segmentation. The cells which are of small size, have budding cells attached to them, lie on the image boundary or are incomplete are removed. These cells are not considered while finding the capsular thickness. Once these cells are removed, the rest of the cells are labelled. Each labelled cell is further enhanced to remove noise and then each cell is put in a separate image. The capsular thickness is then measured

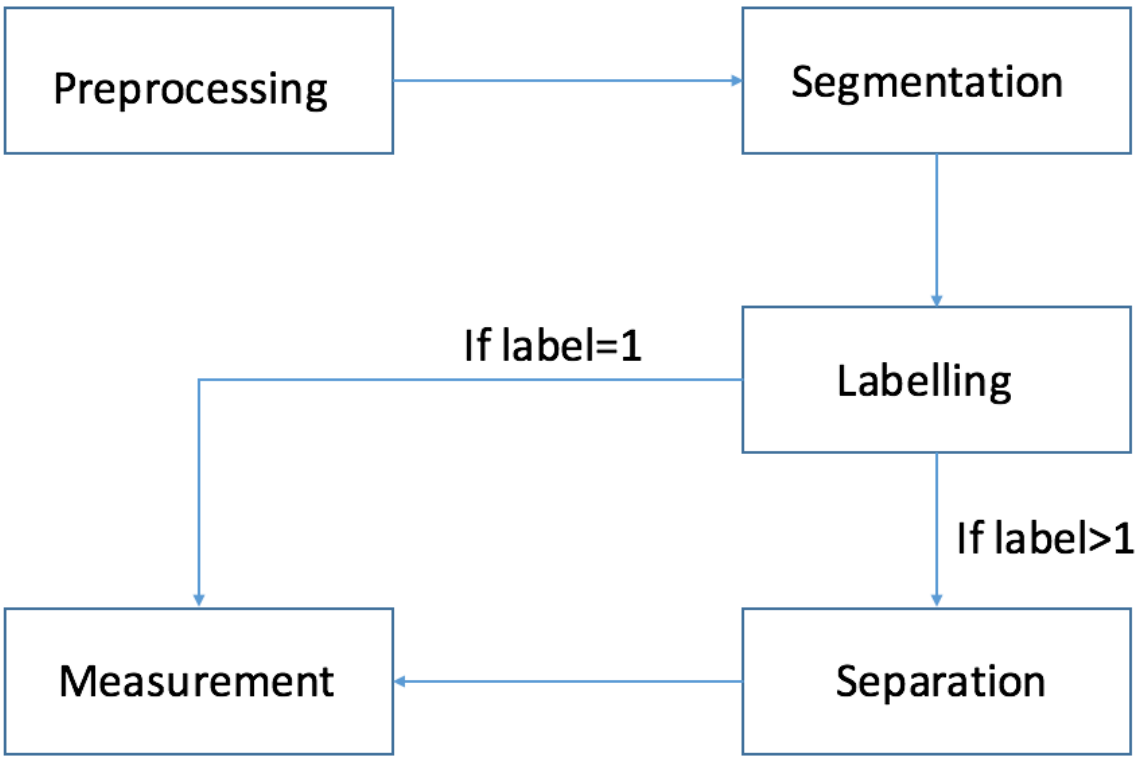


Figure 3.1: The system process flow

## 3.2 PRE-PROCESSING

*STEP 1:* A Gaussian filter was applied which resulted in a binary image of the capsules

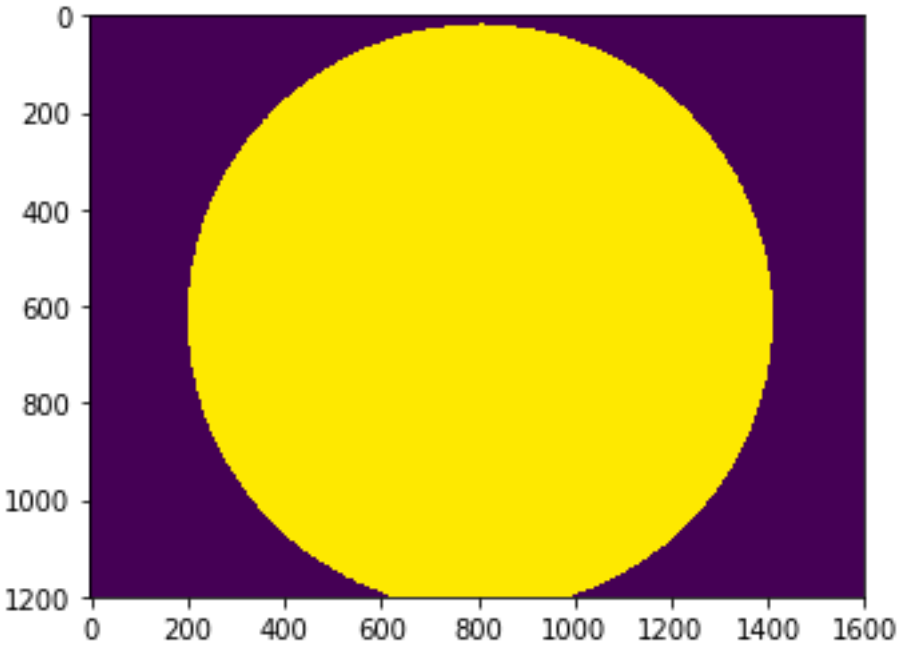


Fig 3.2: Found 1 nuclei

*Step 2:* Detect outer boundary of cell and remove objects that touch the boundary

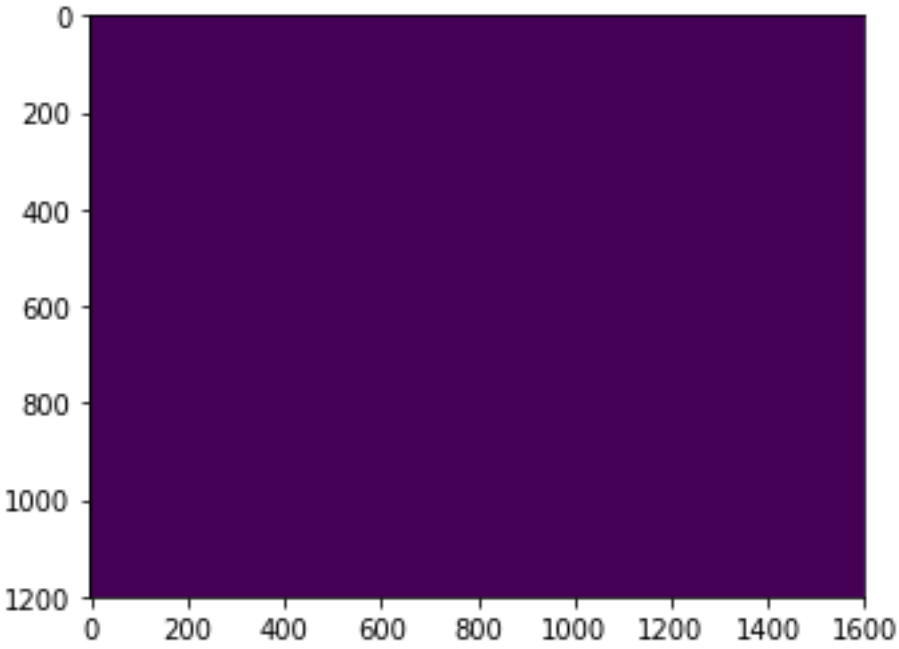


Fig 3.3: After filtering and relabeling, there are 0 nuclei left.

*Step3:* If there are a lot of cells, the cells have a good chance of being cluttered together. The number of cells taken is reduced to half by taking the smaller sized cells.

*Step 4:* Cell touched the boundary and thus was removed. Gaussian blur, a threshold operation and morphological operations were applied on the original image to get the cells in a binary image.

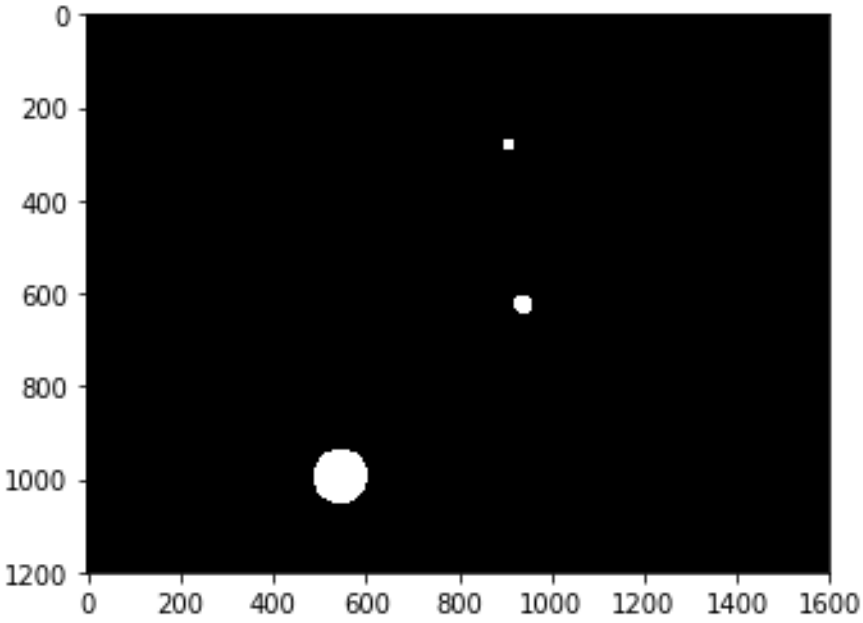


Fig 3.4: Found 3 nuclei

*Step 5:* Cells of small size and which touch the boundary were removed using Mahotas libraries functions remove_bordering and remove_regions.

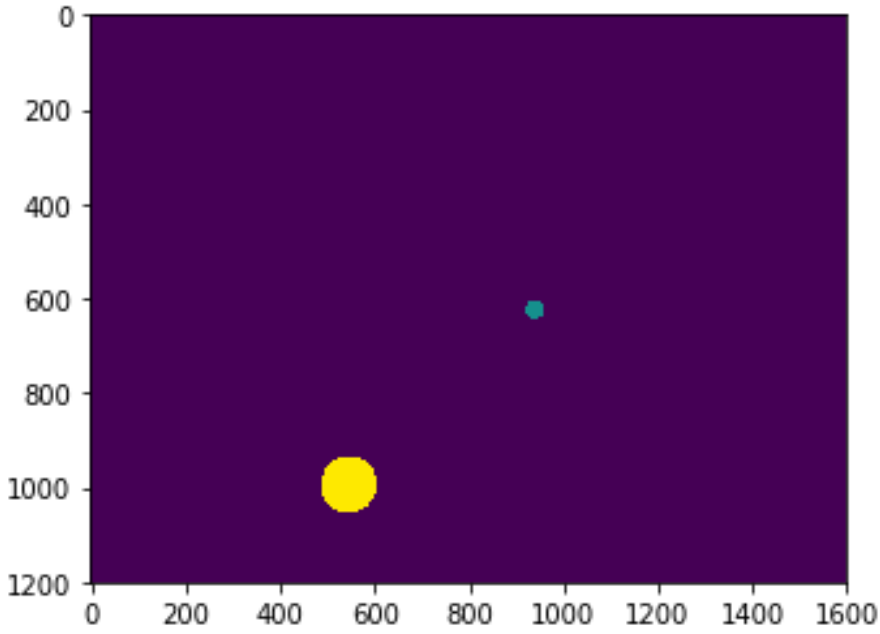


Fig 3.5: After filtering and relabelling, there are 2 nuclei left.

*Step 6:* The image contours were found using findContours function of OpenCV library and then inverted using a bitwise NOT operation.

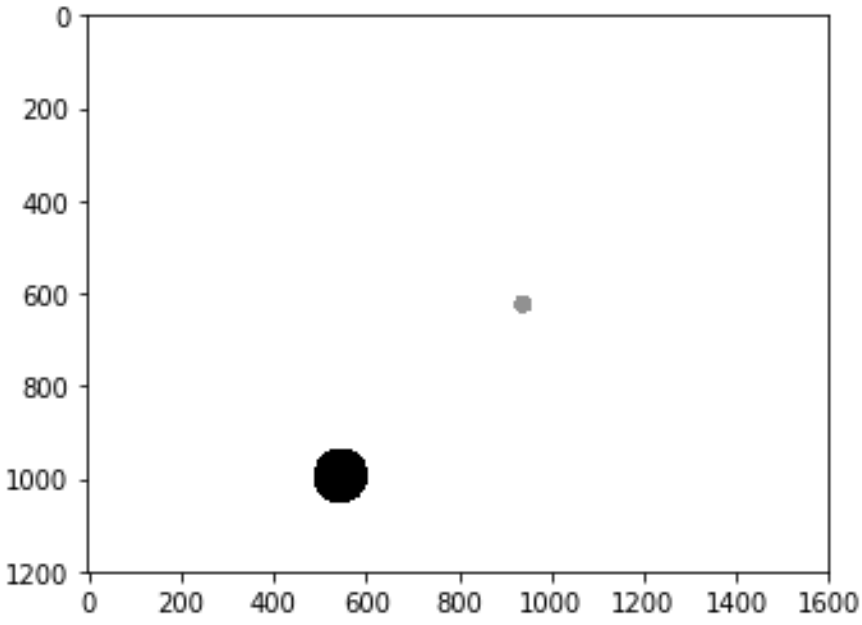


Fig 3.6: Bitwise NOT operation

*Step 7:* The pixel values below a threshold value (here 225 as all images had a white background having pixel values of 225) are extracted from the original image and the rest of the pixels are made white.

Fig 3.7: Background noise removed

*Step 8:* The image is again inverted using the bitwise NOT operation and further enhanced by using the ImageEnhance function from the PIL library to increase the contrast.

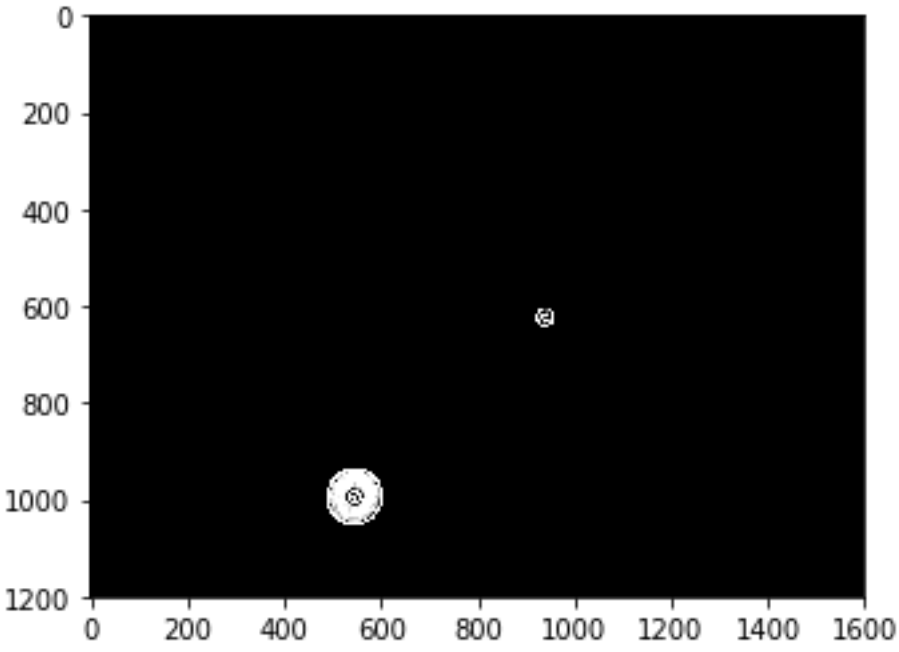


Fig 3.8: Image enhancement

## 3.3 LABELLING

*Step 9:* Performed a connected component analysis on the thresholded image, then initialize a mask to store only the large components. A unique integer is assigned for each blob in label. If the label is zero, then we ignore it as it is the background region. Otherwise, a mask is created for the current label. The number of non-zero pixels are counted and if the blob size is large enough, it is added to the mask. The contours are then found and sorted. For each contour, the minimum enclosing circle is computed and then labelled uniquely.

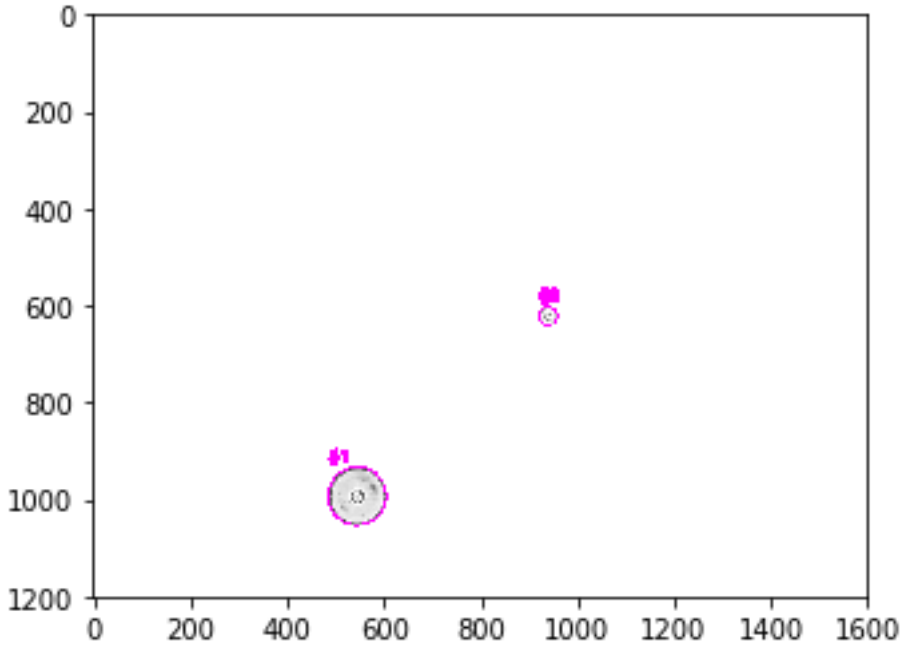


Fig 3.9: 2 Labelled cells

## 3.4 SEPARATION

*Step 10:* The bounding rectangle for each contour is also found by using boundingRect function of OpenCV library to save each cell as a separate image.

Fig 3.10: Cell 1

Fig 3.11: Cell 2

*Step 11:* To each cell, adaptive threshold is performed and the contours are found to detect the outer boundary accurately. The minimum enclosing circle is found and drawn in red as shown below by using the circle function in OpenCV.

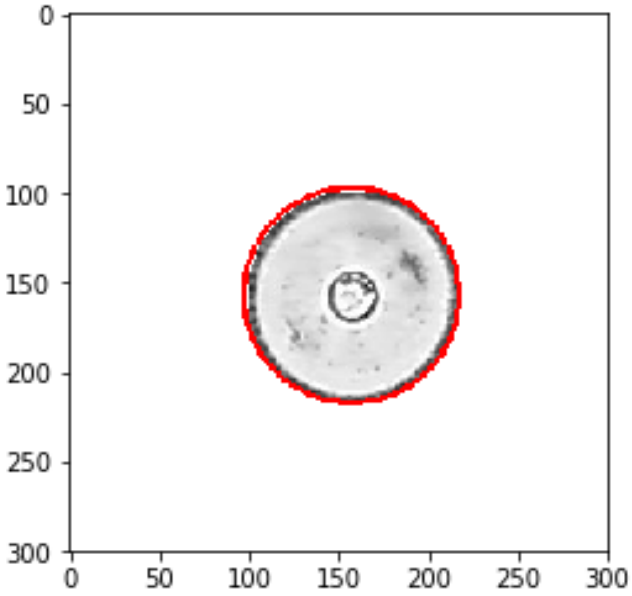


Fig 3.12: Outer boundary shown in red for cell 1

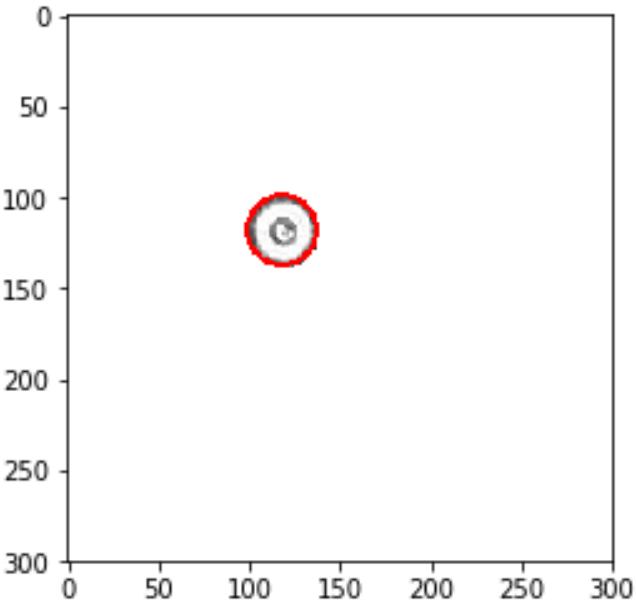


Fig 3.13: Outer boundary shown in red for cell 2

*Step 12:* Threshold operation is again performed and then the outer boundary is made thinner. By using Midpoint circle algorithm for the outer boundary, 8 pixel values in 8 octants on the boundary are stored in a list. On each pixel, the surrounding 8 pixels, and the 16 pixels surrounding those 8 pixels are converted to black.

The threshold value is automatically increased if the radius the the cell is not accurately detected and repeated as long as the threshold value is less than 255.

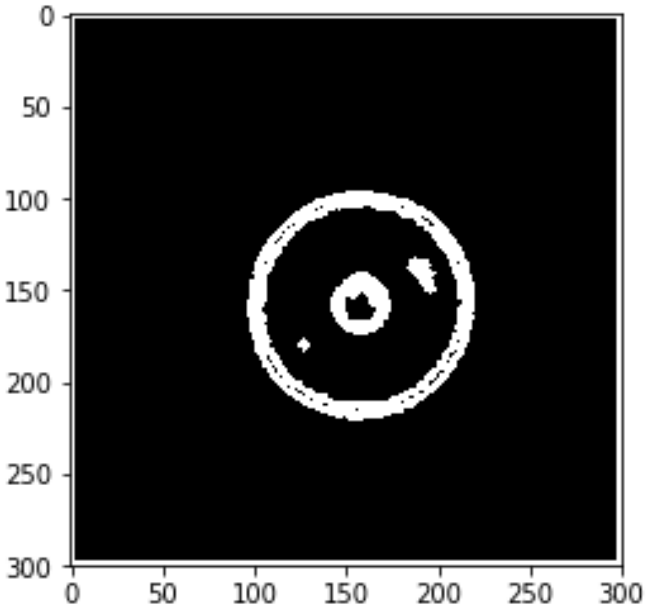


Fig 3.14: Image after threshold operation

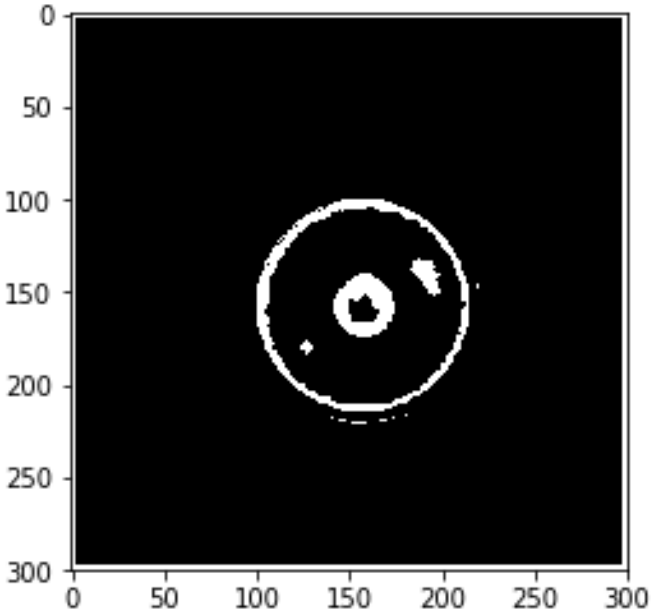


Fig 3.15: Image after boundary thinning

*Step 13:* The inner circle is extracted by removing regions which are outside the area of the outer radius minus 2. Gaussian filter and Otsu thresholding operations are performed again.

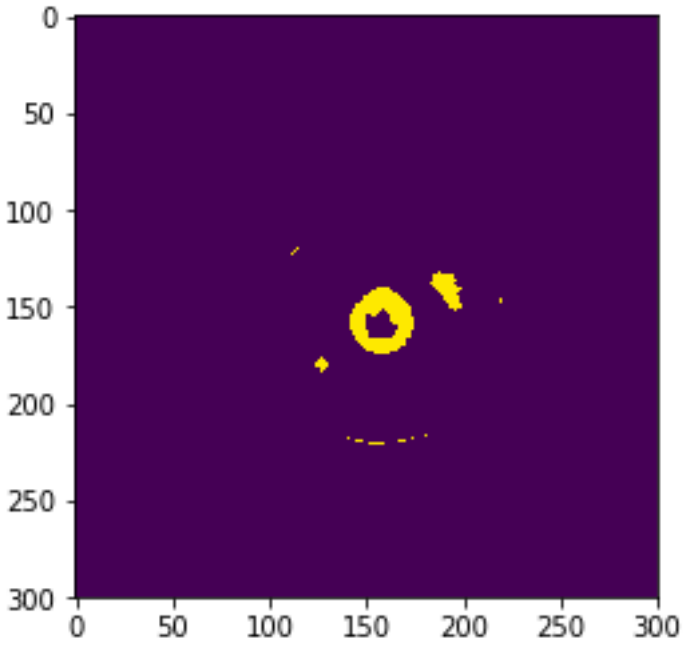


Fig 3.16: After Gaussian filter and Otsu thresholding

*Step 14:* Objects of small size are removed.( Here objects of size less than 200)

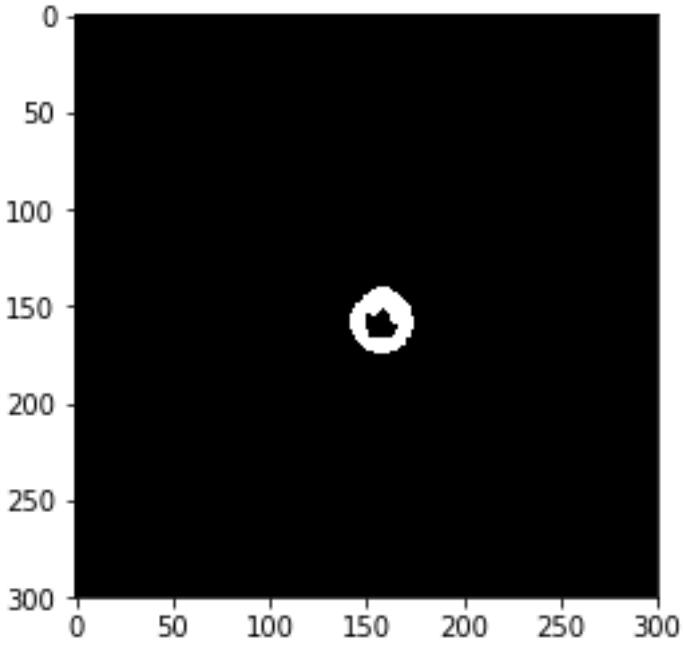


Fig 3.17: Cell without capsular cover extracted

*Step 15:* By applying adaptive thresholding and finding contours, the minimum enclosing circle is found for the cell. Both the detected boundaries are shown below.

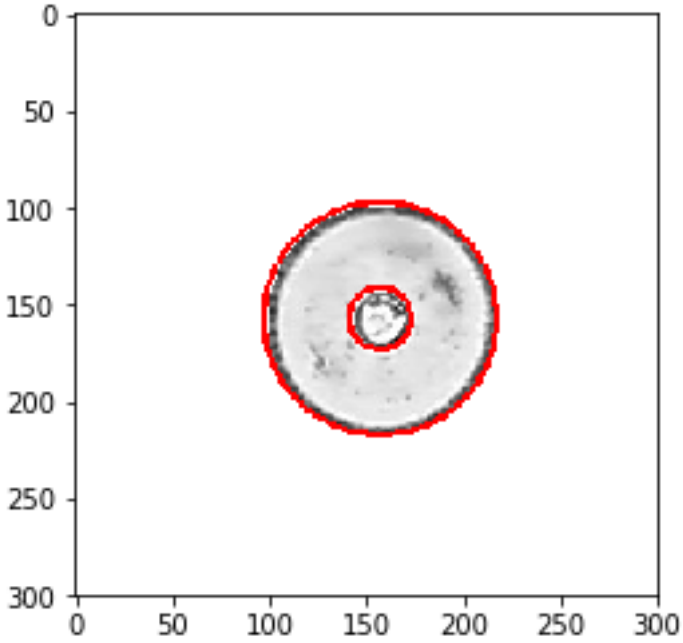


Fig 3.18: Both radius accurately detected

*Step 16:* It was observed that if the cell thickness is below two, the cell is either not clear or it is not a cell. It could be a white patch as illustrated below. Such a cell is labelled as incorrect in the results. The measurements for such a cell is also shown below.

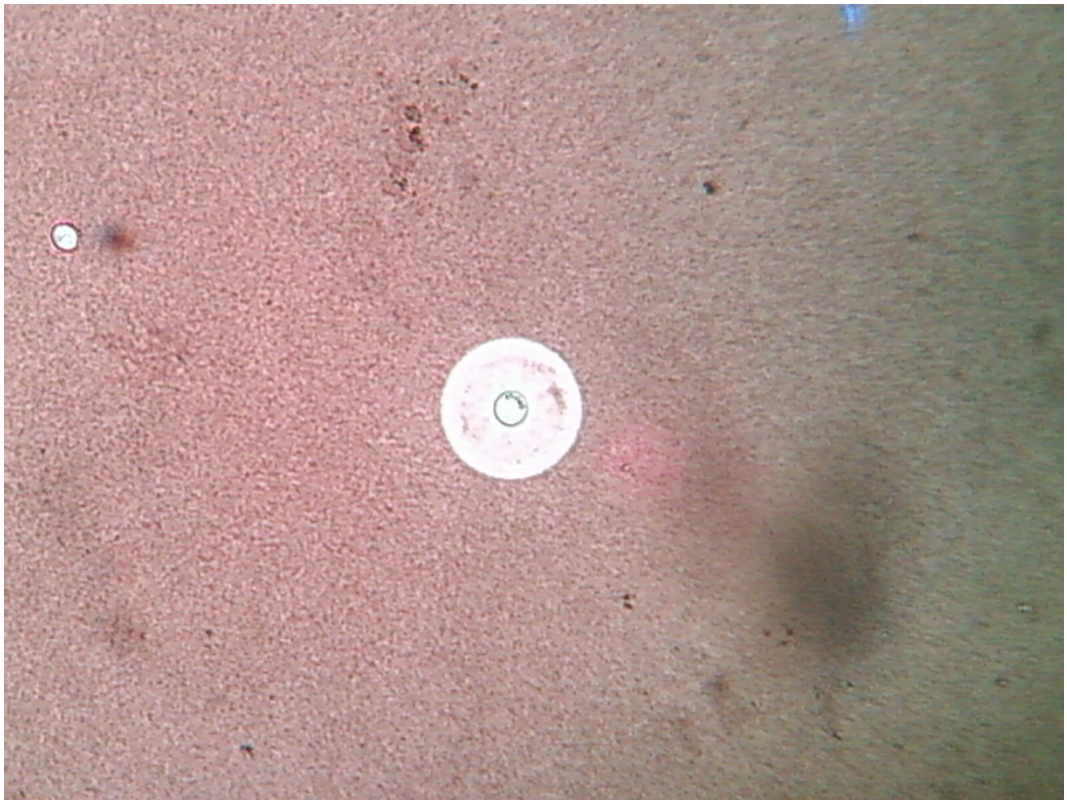

Fig 3.19 Image with a white patch

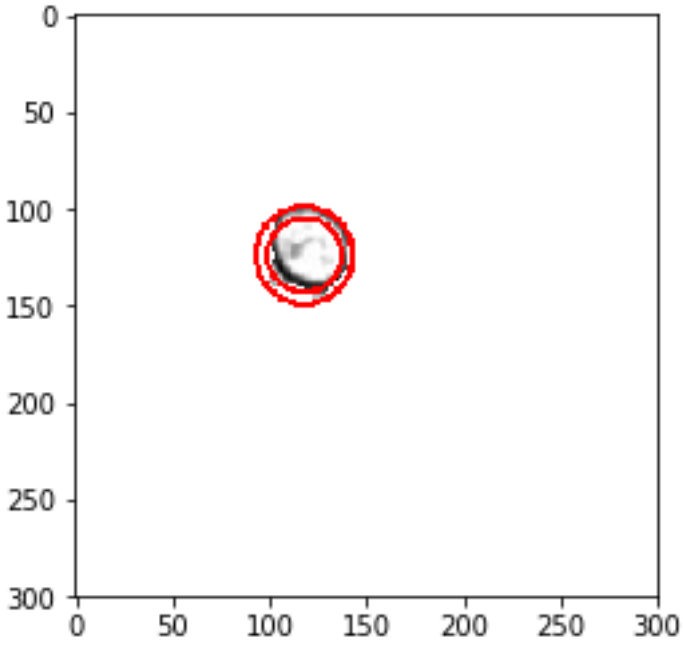


Fig 3.20: Cell with no outer capsule or is not clear.

## 3.4 TOOLS USED

Python Libraries used:

- Numpy
- Matplotlib
- Pandas
- Sklearn
- PIL
- OpenCV
- Pylab
- Skimage
- Scipy
- Mahotas

# CHAPTER 5
# RESULT ANALYSIS

## 4.1 INTRODUCTION

All the above tests are conducted on a set of images of the yeast cell Cryptococcus taken from the department of microbiology in KMC. The test results show that the capsular thickness of each cell was accurately detected for all the images. The results have been taken only for single cells. Multiple and budding cells have not been considered.

## 4.2 RESULT ANALYSIS

The results were stored in a pandas dataframe. Thickness is the capsular thickness, inner is the cell radius and total is the overall radius of the cell with the capsular cover.

Table 4.1: Final results for each cell

| | thickness | inner | total |
|---|---|---|---|
| 1 | 39.120459 | 21.249334 | 60.369793 |
| 2 | 6.941704 | 13.053993 | 19.995697 |

The results can be easily observed in the graph below for each cell.

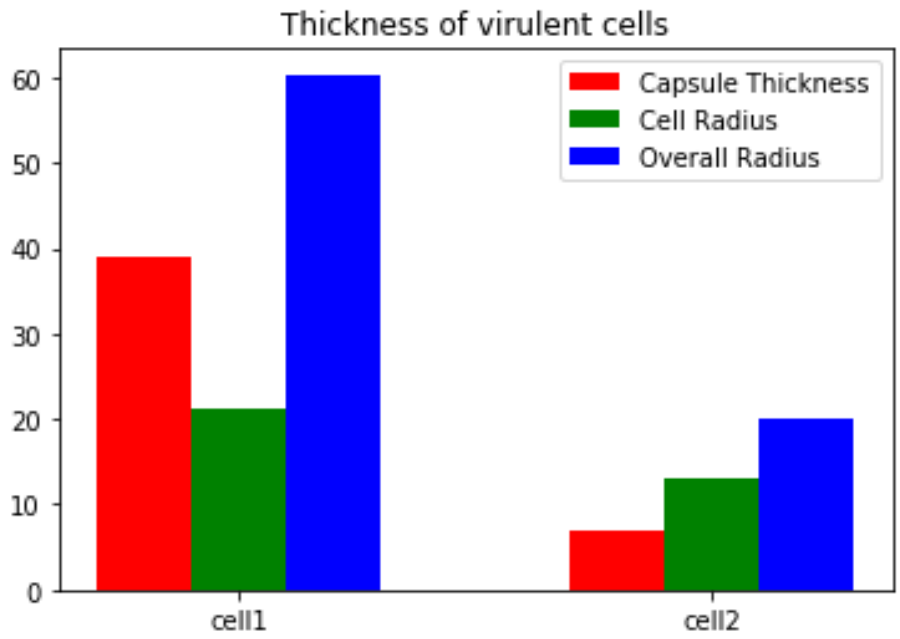


Fig : 4.1: Graph showing results

For cells which are not clear, they are marked below as incorrect to distinguish them.

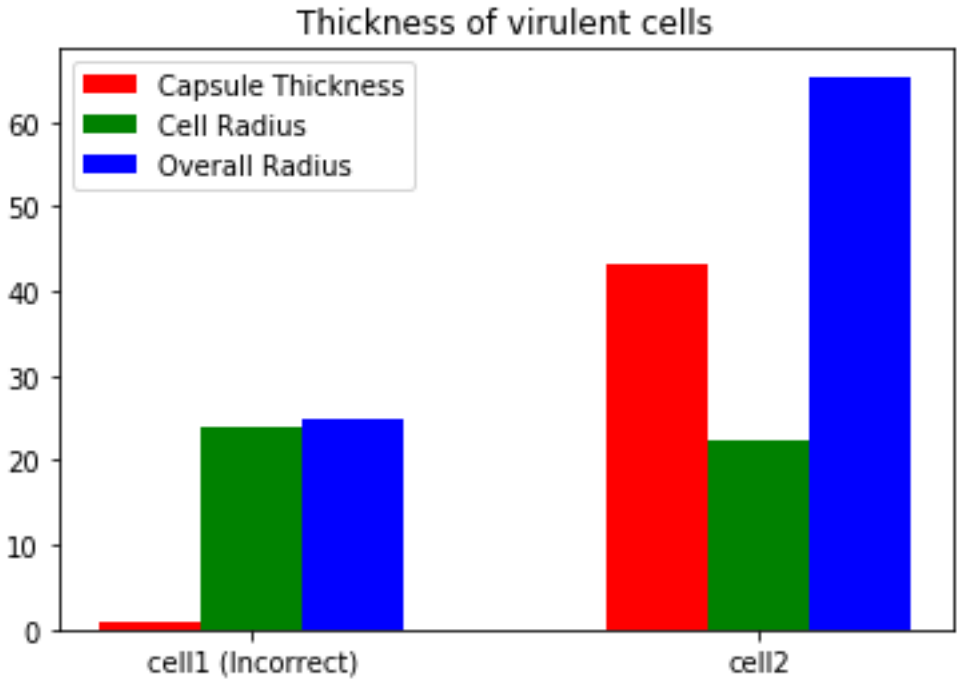


Fig 4.2: Graph showing results of unclear cells also

A Graphical user interface is created to display the results. The dataframe, number of cells in the image and graph are displayed.


Fig 4.3: Graphical user interface

Each step is also displayed to help understand the process of measurement.

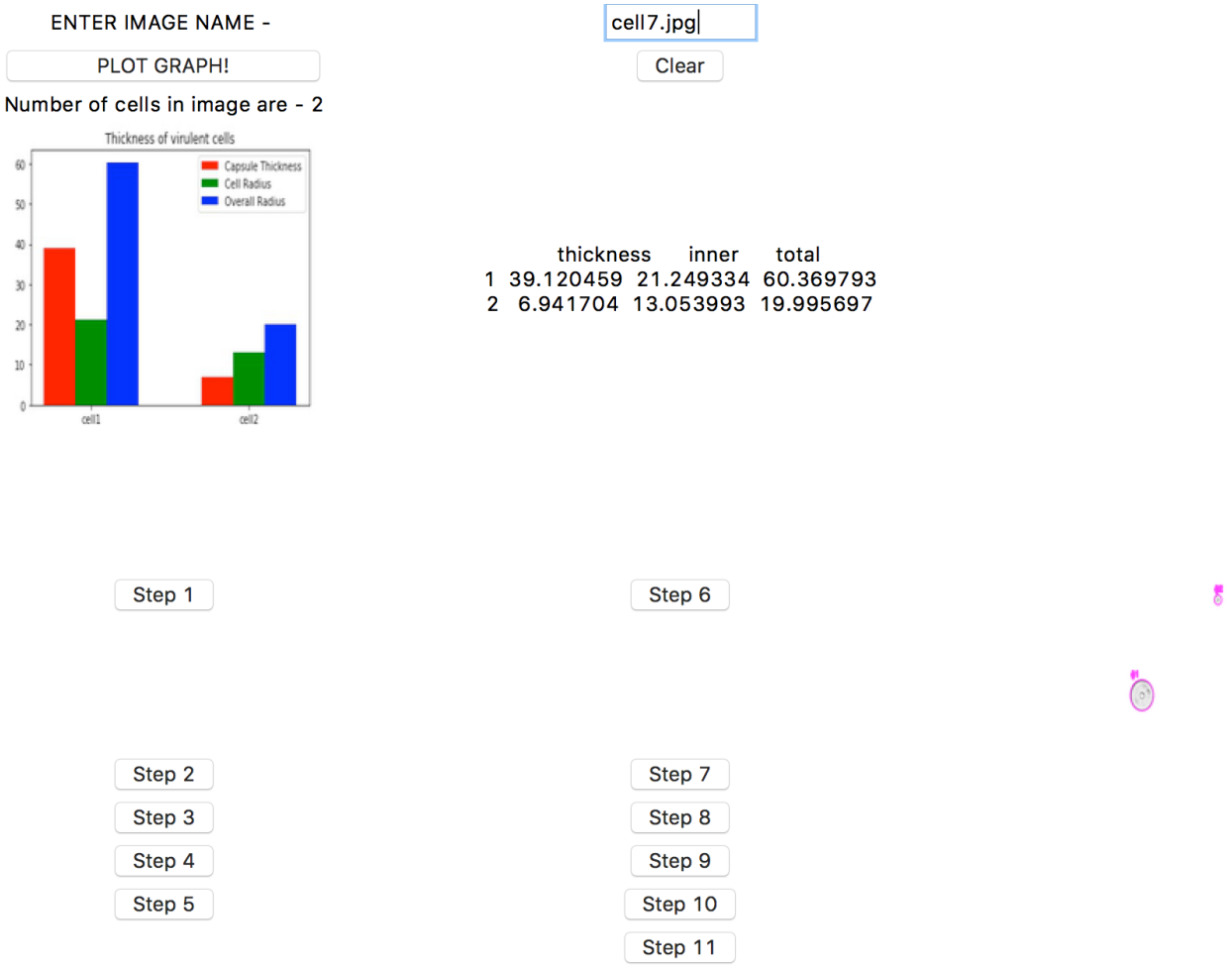


Fig 4.4: Steps involved in measurement

# CHAPTER 6
# CONCLUSION AND FUTURE SCOPE

## 5.1 SUMMARY

The objective of this project was to study the yeast cell Cryptococcus and design an efficient measuring system to find the thickness of the capsular region around the cell which contributes to its virulence. It made use of python libraries to pre-process the images and display the results on a graphical user interface using Tkinter.

## 5.2 CONCLUSION

Previous studies show that capsule thickness and cell area was manually measured and no automated procedure existed. The goal is to create a tool which can find the capsule thickness for any cell in an image with single cells to detect pathogen conditions. This will be of importance to research in this direction.

## 5.3 FUTURE SCOPE OF WORK

The area of study that this project belongs to, is still in its infancy stage and there is immense Scope for future work. In an increasingly health conscious world driven by technology, there is a lot of scope for further application and refinement in this project. The future scope of this project has been discussed as below:

- The measurement system includes only single cell images. Multiple cells can also be used in the future to widen the scope of the research.
- The biggest problem faced was that the cells were not clear and had to be heavily pre-processed. Also a lot cells were cluttered together and could not be distinguished from one another. This area could be researched more to refine the project.
- Other factors beyond features in the image such as rate of growth of the capsular cover could also be considered to make a classification system.

# REFERENCES

## *Journal / Conference Papers*


[1] Yu, Bo Yang, et al. "Image processing and classification algorithm for yeast cell morphology in a microfluidic chip." *Journal of biomedical optics* 16.6 (2011): 066008-066008.

[2] van der Putten, Peter, et al. "Classification of yeast cells from image features to evaluate pathogen conditions." *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007.

[3] Liu, Jinshuo, et al. "Detecting Virulent Cells of Cryptococcus Neoformans Yeast: Clustering Experiments." *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 1. IEEE, 2006.

[4] O. Zaragoza, A. Casadevall, "Experimental modulation of capsule size in Cryptococcus neoformans," Biological Procedures Online • Vol. 6 No. 1 March 3, 2004, www.biologicalprocedures.com

[5] M. Ohtani, A. Saka, F. Sano, Y. Ohya, and S. Morishita, "Development of image processing program for yeast cell morphology," J. Bioinf. Comput. Biol. 1(4), 695–709 (2004).

[6] Niemistö, Antti, et al. "Computational methods for estimation of cell cycle phase distributions of yeast cells." *EURASIP Journal on Bioinformatics and Systems Biology* 2007 (2007): 2-2.

[7] R. Holte, "Very simple classification rules perform well on most commonly used datasets", Machine Learning 11, 63–91 (1993).

[8] Dykstra, MARK A., L. O. R. R. A. I. N. E. Friedman, and J. W. Murphy. "Capsule size of Cryptococcus neoformans: control and relationship to virulence." *Infection and immunity* 16.1 (1977): 129-135.

[9] O. Zaragoza, A. Casadevall, "Experimental modulation of capsule size in Cryptococcus neoformans," Biological Procedures Online • Vol. 6 No. 1 March 3, 2004, www.biologicalprocedures.com

[10] G. Janbon, "Cryptococcus neoformans capsule biosynthesis and regulation," FEMS Yeast Res 4, 765-71 (2004)

[11] M. Ohtani, A. Saka, F. Sano, Y. Ohya, and S. Morishita, "Development of image processing program for yeast cell morphology," J. Bioinf. Comput. Biol. 1(4), 695–709 (2004)

[12] D. L. Pham, C. Y. Xu, and J. L. Prince, "Current methods in medical image segmentation," Annu. Rev. Biomed. Eng. 2, 315–337 (2000)

[13] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., p. 954, Pearson/Prentice Hall, Upper Saddle River, NJ (2008)

[14] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson, "Integrative analysis of cell cycle control in budding yeast," Mol. Biol. Cell 15(8), 3841–3862 (2004).

# PROJECT DETAILS

| | | | |
|---|---|---|---|
| *Student Details* | | | |
| **Student Name** | **Arushi Gupta** | | |
| Register Number | 130905372 | Section / Roll No | B / 27 |
| Email Address | Arushigupta148@gmail.com | Phone No (M) | 9663557206 |
| | | | |
| *Project Details* | | | |
| **Project Title** | Classification of Yeast Cell Cryptococcus to detect virulent cells | | |
| Project Duration | 4 months | Date of reporting | 5th may 2017 |
| | | | |
| *Organization Details* | | | |
| **Organization Name** | Manipal Institute of Technology | | |
| Full postal address with pin code | Udupi Karkala Road, Manipal, Karnataka 576104 | | |
| Website address | https://manipal.edu/mit.html | | |
| | | | |
| *External Guide Details* | | | |
| **Name of the Guide** | | | |
| Designation | | | |
| Full contact address with pin code | | | |
| Email address | | Phone No (M) | |
| | | | |
| *Internal Guide Details* | | | |
| **Faculty Name** | **Vidya V.** | | |
| Full contact address with pin code | Dept of Computer Science & Engg, Manipal Institute of Technology, Manipal – 576 104 (Karnataka State), INDIA | | |
| Email address | Vidya.prakash@manipal.edu | | |