Cancer Detection using Image Processing Techniques Based on Cell Counting, Cell Area Measurement and Clump Detection

BRAC
UNIVERSITY

Inspiring Excellence

Monirul Islam Pavel | **14301141**
Ashique Mohammad Sadique | **12201103**
Rifat Ahmed Ritul | **13101205**
SynthiaKhan | **13101006**
SnigdhaNath | **14101125**

Department of Computer Science and Engineering
BRAC University

Supervised by Dr. Md. Ashraful Alam
Submitted on: 21 August 2017

DECLARATION

We, hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This Thesis, neither in whole or in part, has been previously submitted for any degree.

Signature of Supervisor:

Dr. Md. AshrafulAlam

Signature of Authors:

Monirul Islam Pavel

Ashique Mohammad Sadique

Rifat Ahmed Ritul

Synthia Khan

SnigdhaNath

Acknowledgement

Firstly we would like to thank the almighty for enabling us to initiate our research, to put our best efforts and successfully conclude it.

Secondly we offer our sincere and heartiest gratitude to our respected Supervisor Dr. Md. AshrafulAlam for his contribution, guidance and support in conducting the research and preparation of the report. His involvement & guidance has been of tremendous value throughout our research.

Last but not the least, we are thankful to the faculties, seniors, friends and our family who has motivated and inspired us throughout this journey. We would also like to acknowledge the assistance we received from numerous resources over the Internet especially from fellow researchers work.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
DNA	Deoxyribonucleic acid
NMR	Nuclear magnetic resonance
RGB	Red, Green, Blue

Abstract

We proposed cancer cell detection using image processing techniques based on cancer cell counting, cell area measurement and clump detection. The proposed system autonomously detects traits of cancer in microscopy images of biopsy samples. Previous similar attempts lack flexibility in terms of features and the level of accuracy is not consistent on respective type of cancer. The system pre-processes the input image by means of gray scaling, binarization, inversion, median filtering and flood fill operation. The pre-processed image then undergoes "cell counting", "area measurement" or "clump detection" depending on the type of trait to be detected. Several sets of images were processed using this methodology and the system was fine-tuned using the feedback from these test runs. The proposed method can be effectively used for autonomous cancer cell detection, which will significantly accelerate cancer cell researches and open up new dimensions.

Chapter 1
Introduction

1.1 Motivation:

The term cancer is referred to obstreperous division of abnormal cells. Cancer cells can spread to other parts of the body through the blood or lymphatic systems and create tumors. But it must be noted that all tumors are not cancerous. Tumors can be benign (not cancerous), or malignant (cancerous). There are over one hundred types of cancer that has been recognized and each type has many subtypes which has variations of their own. This immense variation makes cancer detection very complex, especially in the preliminary stages. The causes of cancer, in most cases, are still not very well understood. Hence treating cancer becomes even more strenuous.

And due to this enormous complexity of the disease scientists, doctors and engineers all over the over the world are researching on the field of cancer to achieve a better understanding of cancer and find absolute cures for each type of cancer in the process. Even though the process is lengthy and difficult, but knowing more will enable doctors to cure cancer patients more effectively.

This motivated us to think about the mechanism of cancer detection and use technology to speed up the process. If cancer researchers are able to automatically detect cancer cells via means of image processing, this can save tremendous amounts of time and also increase efficiency of the research, since the human error factor will cease completely.

1.2 Literature Review

Kumar et al. approached a cancer detection system using clinical significant and biologically interpretablefeatures. They performed K-mean cluster and then ROI segmented by texture based segmentation. The various stages involved in the proposed methodology include enhancement of microscopic images, segmentation of background cells, features extraction, and finally the classification. For the enhancement of the microscopic biopsy images, the contrast limited adaptive histogram equalization approach is used and for the segmentation of background cells -

means segmentation algorithm is used. In feature extraction phase, various biologically interpretable and clinically significant shape and morphology based features are extracted from the segmented images which include gray level texture features, color based features, color gray level texture features, Law's Texture Energy (LTE) based features, Tamura's features, and wavelet features. Finally, the -nearest neighborhood (KNN), fuzzy NN, and support vector machine (SVM) based classifiers are examined for classifying the normal and cancerous biopsy images. These approaches are tested on four fundamental tissues (connective, epithelial, muscular, and nervous) of randomly selected 1000 microscopic biopsy images. Finally, the performances of the classifiers are evaluated using well known parameters and from results and analysis, it is observed that the fuzzy KNN based classifier is performing better for the selected features set [1].

Patil,G.Bhagyashri. , Jain, Sanjeev.N. approaches some pre-processing methods while detecting lung cancer such as noise removal and enhancement ,author also gave proper idea of some post-processing methods such as thresholding , marker-controlled watershed segmentation, feature extractions, binarizationetc [2].

Ramin M., Ahmadvand, P., Sepas-Moghaddam, A.Dehshibi, M.M. used image analysis technique for counting number of cells in Immunocytochemical (ICC) images [3]. The proposed system contains four major steps: Pre-processing, Classification, Separating Bound Nucleus and Cell Counting. Pre-processing consists of removal of random noise by smoothing spatial filter. Morphological open operator is utilized to eliminate images' background. Banding noise is removed by subtracting median of the red channel from all channels. In order to separate nucleus from antigens, nearest neighbor classification method with Euclidean distance metric is used in L*a*b color space. The bound nucleus is separated by local thresholding algorithm. For this purpose, statistical analysis is done and optimal threshold is found with the help of genetic algorithm. Finally, cell counting is done by tracing the boundaries. From the results, the Error Ratio and Standard Deviation of the proposed method are 6.75% and 6.39% respectively [3].

In the paper named "Automated counting of mammalian cell colonies by means of a flatbed scanner and image processing", the authors counted blood cancer cells by developing a device which employs a flatbed scanner to image 12 60-mm petri dishes at a time. Two major problems in automated colony counting are the clustering of colonies and edge effects. The resulting system was compared with two manual colony counts, as well as with automated counts with the Oxford OptronixColCount colony counter for cell lines V79 and HaCaT after using standard image analysis and implementing an inflection point algorithm [4].

Venkatalaksmi B. and Thilagavathi K. presented a method for automatic red blood cell counting using hough transform. The algorithm for estimating the red blood cells consists of five major steps: input image acquisition, pre-processing, segmentation, feature extraction and counting. In pre-processing step, original blood smear is converted into HSV image. As Saturation image clearly shows the bright components, it is further used for analysis. First step of segmentation is to find out lower and upper threshold from histogram information. Saturation image is then divided into two binary images based on this information. Morphological area closing is applied to lower pixel value image and morphological dilation and area closing is applied to higher pixel value image. Morphological XOR operation is applied to two binary images and circular hough transform is applied to extract RBCs [5].

Bergmeir et al. proposed a model to extract the texture features by using local histograms and GLCM. The quasisupervised learning algorithm operates on two datasets, the first one having normal tissues labeled only indirectly and the second one containing an unlabeled collection of mixed samples of both normal and cancer tissues. This method was applied on the dataset of 22,080 vectors with reduced dimensionality, 119 from 132. The regions having the cancerous tissues were accurately identified having true positive rate 88% and false positive rate 19%, respectively, by using manual ground truth dataset [6].

Carlos A. B. Mello et al. presented two methods for mosquito eggs counting. These methods are based on a different color model. In the first method, RGB image is converted into HSL color model (Hue, saturation, Lightness). From these three components, the hue image is extracted as it contains information about color tone. Huang thresholding algorithm is applied to the hue image for binarization. A connected components algorithm is used to label the connected regions of the image. Filtering is done using morphological opening operation with structuring element defined in the form of egg. At the last step, it is considered that egg occupies area of 170 pixels. The number of eggs is calculated by dividing the total amount of white pixels by this average area. The second method is based on converting RGB sub-image to YIQ one. From these components, I band is segmented in two ways: by using linearization with fix threshold of 200 and by binarization using k-means clustering method. For performing egg counting in this method, it is considered that the average size of mosquito egg is 220 pixels [7].

1.3 Thesis Orientation

The rest of the thesis is organized as follows:

- Chapter 2 discusses the fundamentals of image processing.

- Chapter 3 presents the proposed model of our research

- Chapter 4 demonstrates the results found in our research

- Chapter 5concludes the thesis and states the future research directions.

Chapter 2
Fundamentals of Biomedical ImageProcessing

2.1 Image Processing

Image processing was first used in the early 1920s in a paper industry where images were coded for a submarine cable transfer & reconstructed by a telegraph printer at the receiving point. In the mid to late 1920s, there had been improvements in the system. In 1964, image processing was used to improve the images of the moon taken by the Ranger 7 space probe. Such techniques were used in the other space missions as well. In the 1970s, image processing began to be used in the fields of medical science. In 1979, Allan M. Cormack and Godfrey N. Hounsfield jointly received The Nobel Prize for the invention of computer assisted tomography. Nowadays digital image processing is getting more and more attention because of the focus on two principal areas [9]:

- Improvements in the image information for human interpretation
- Processing of image for autonomous perception

Uses of image processing are expanding and now they are used in almost all areas.
Fields that use Image Processing:

- Medical Science
- Robot Vision
- Pattern Recognition
- Restorations & Enhancements
- Remote Sensing
- Transmission & encoding
- Digital Cinema

Visual representation of any two or three dimensional scene is called an image. To be precise, image is an array or a matrix where pixels (image elements) are arranged in a columns and rows. It can be considered as a two dimensional function *f(x, y)* where *x, y* are two spatial coordinates that represents the intensity or the gray level of the image at that point.Where there is image, image processing is involved. Image processing is one of the most widely used & rapidly growing technologies. It is the process of analyzing and manipulating a digitized image for quality enhancement or to extract any information from it.Image processing basically follows these steps [10]:

- Importing image via image acquisition tools
- Analyzing and manipulation of image
- Output can be an image or report based on image analysis

Two types of methods are used for image processing named: Analogue and Digital Image processing. Analogue image processing can be used for hard copies like printouts and photographs since analogue are required for human viewing. Digital image processing involves manipulation of images in digitized way using a computer or other devices. There are three phases images go through for digital image processing [9]:

- Low-level image processing
- Mid-level image processing
- High-level image processing

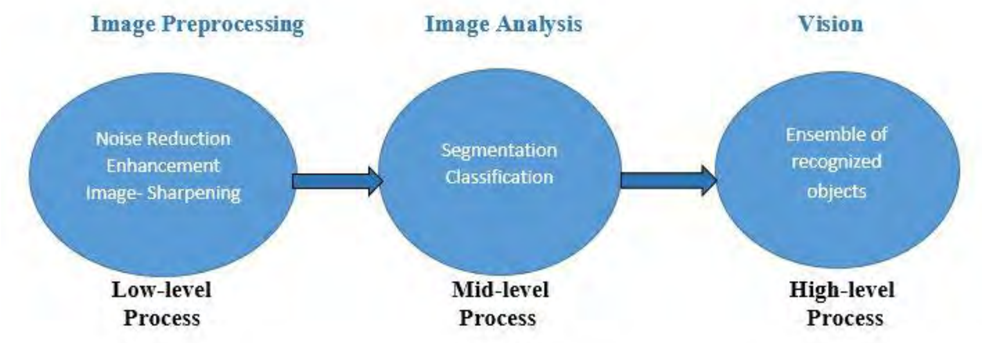


Fig. 2.1.1 Phases of Digital Image Processing

2.2 Stages of Digital Image Processing.

In digital image processing there are a few key stages that are needed to be followed. The stages are as followed:

1. Image Acquisition
2. Image Enhancement
3. Image Restoration
4. Morphological Processing
5. Segmentation
6. Object Recognition
7. Representation & description

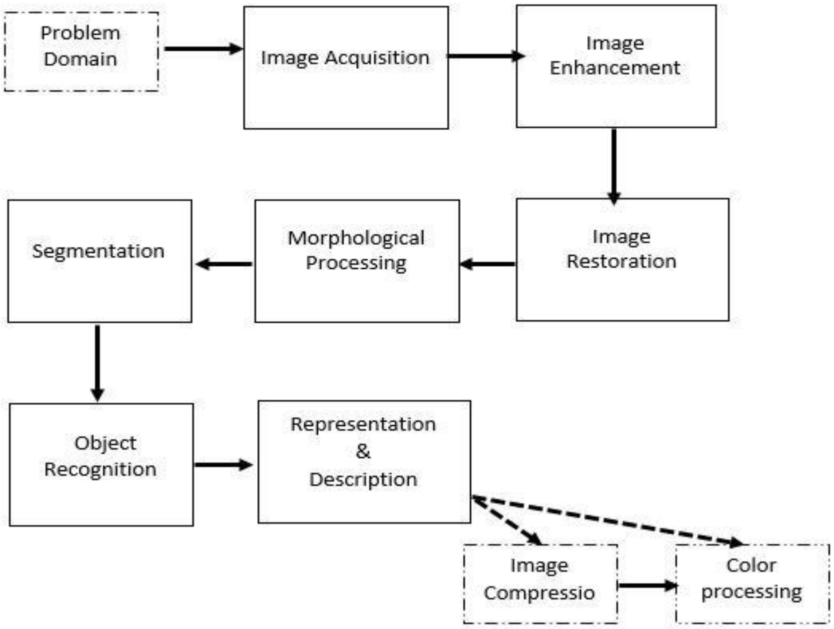


Figure 2.2.1 Architecture of Fundamental Image Processing

2.2.1 Image Acquisition

Illumination of a scene and the by absorbing energy reflected by the light in the scene is how image generates. Sensors turn energy that's absorbed from illumination into digital image. ImageAcquisition is the first stage. It refers to the preprocessing of an image such as: scaling [9].

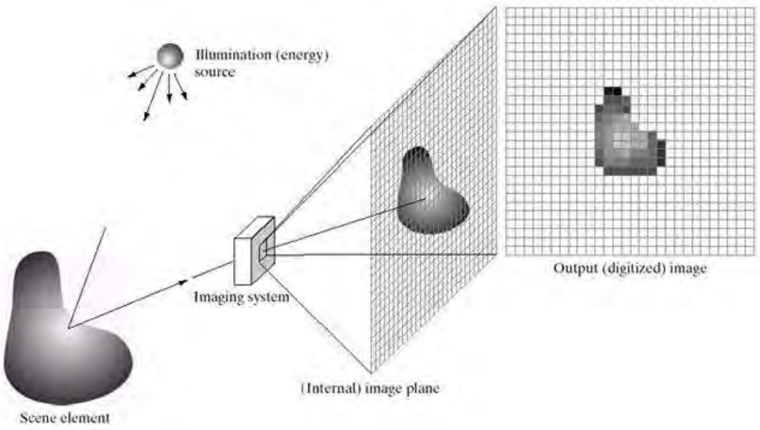
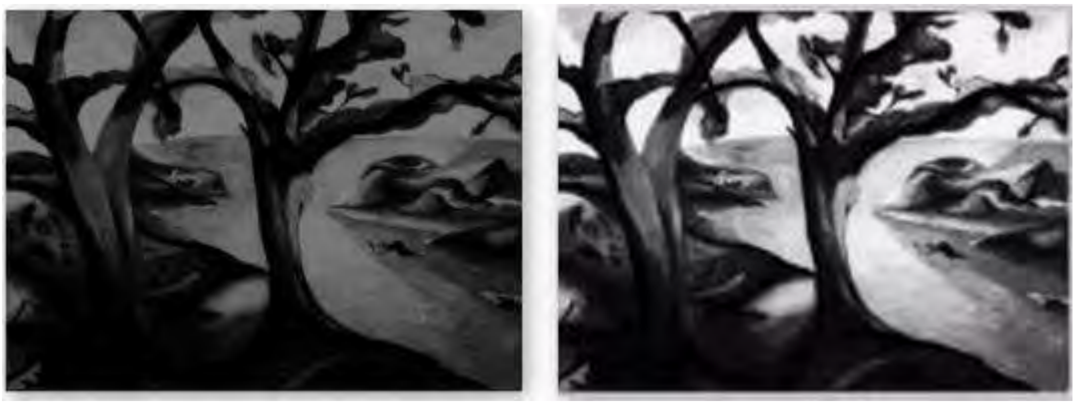


Fig. 2.2.2 Process of image acquisition

2.2.2 Image Enhancement

Image enhancement is improving the interpretation of an image for better human view. Basically it means improving an image's quality by manipulating it in terms of brightness, blurring, contrast, saturation etc. so that the resulting image is more suitable for a particular application. There is no particular set of rules to that a proper enhanced image. Viewer is the ultimate judge of it. Two types of methods are followed for image enhancement. [10] They are:

- Spatial domain method, operates directly on pixels
- Frequency domain method, operates on Fourier transformation of an image

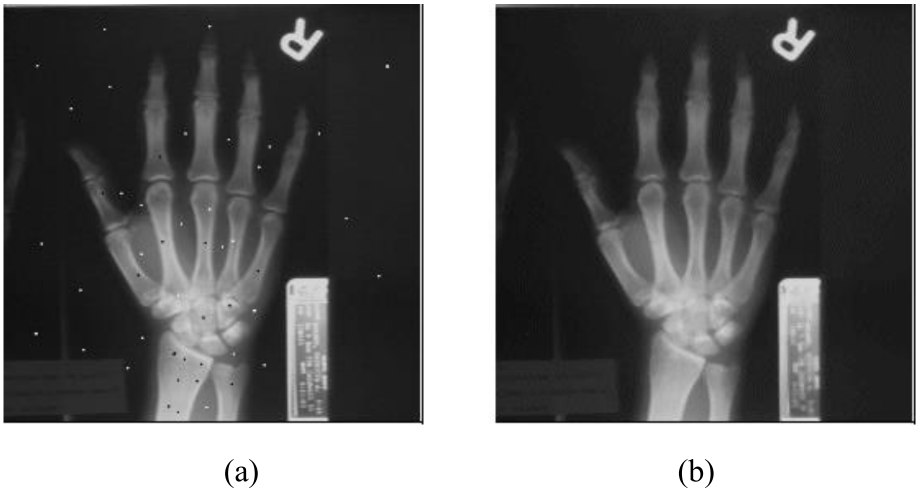


(a) (b)

Fig. 2.2.3 (a) Host Image (b) Enhanced Image

2.2.3 Image Restoration

Image restoration is somewhat similar to image enhancement although it is an objective process whereas image enhancement is a subjective process. Image restoration uses the prior knowledge of degradation and uses it to retrieve an image. Often it is seen that original or raw images are not always in the best quality. Image restoration recovers the original image by applying reverse approach. Examples of image restoration are: noise reduction, image sharpening [9].



(a) (b)

Fig. 2.2.4 (a) Host Image (b) Image after restoration with median filter

2.2.4 Morphological Processing

Morphological image processing is a set of nonlinear operations related to the extracting and describing morphology (shape,form, structure) of an image. It refers to a mathematical tool that investigates the geometric structure of binary and gray scale image. It is especially suitable for binary images. Although it can be applied in gray scale images in a way that their light transfer functions are unknown and their absolute pixel values are of no or minor interest. It is based on set theory.

Morphological image processing probes an image with structural element which is positioned in all possible locations of an image. These elements are compared with their neighborhood. Some of the elements check whether it fits in the neighborhood whereas others examine whether they intersect with the neighborhood or not.

Goal of this image processing is to differentiate shape information from an irrelevant one. Erosion and Dilation are the most elementary operations of morphological image processing.

The basic morphological algorithms that we have used for our research purposes are[11, 12]:

- Boundary Extraction
- Region Filling
- Convex hull
- Pruning
- Extraction of Connected Components

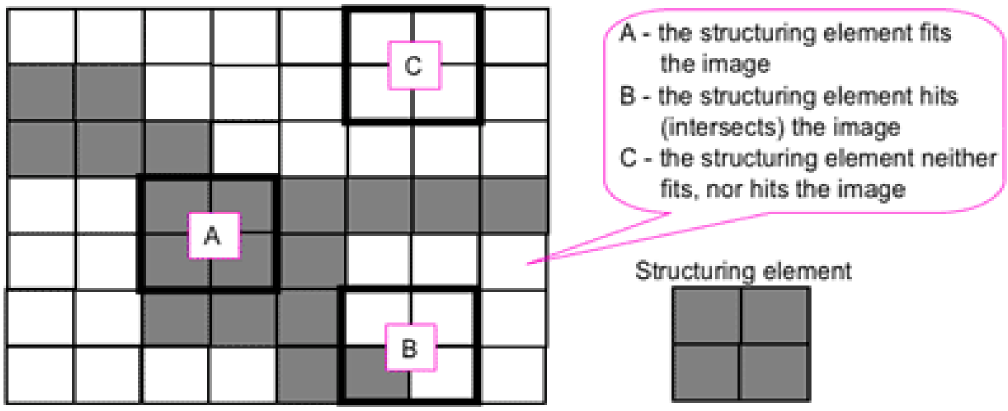

C
A
B
A - the structuring element fits the image
B - the structuring element hits (intersects) the image
C - the structuring element neither fits, nor hits the image
Structuring element

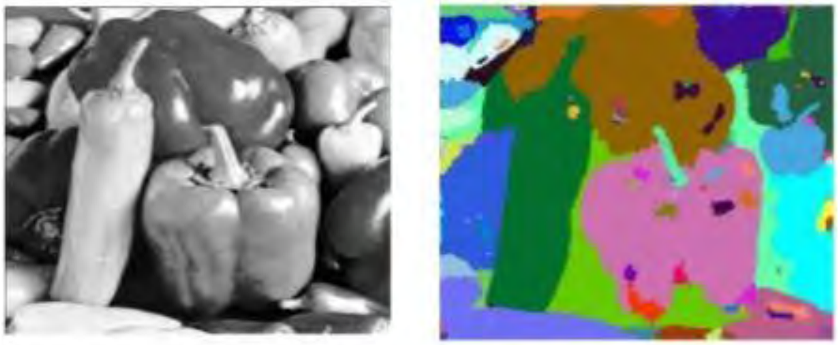
Fig. 2.2.5 Probing an image with structuring elements(white and gray pixels have zero and nonzero values)

2.2.5 Segmentation

Segmentation involves a procedure that partitions or breaks an image apart into its constituents. It alludes to another progression in image processing technique where inputs are pictures and outputs are attributes extracted from pictures. It divides images into its constituents regions or objects. Its accuracy level determines the success or failure of the analysis procedure. Segmentation is works withmeasurements taken from an image that might be gray level, color, texture, depth or motion. Segmentation algorithms are generally based on the two basic properties of intensity values[13]:

- Discontinuity:Partition of an image based on sharp changes in intensity (e.g. edges in an image).
- Similarity:Partition of image into regions that are similar according to a set of predefined criteria; which includes thresholding, region growing, region splitting and merging.

Example 1: Segmentation based on greyscale:



(a) (b)

Fig 2.2.6 (a) Host Image (b) Segmentation based on greyscale

Example 2: Segmentation based on texture, enables object surfaces with varying patterns of grey to be segmented:

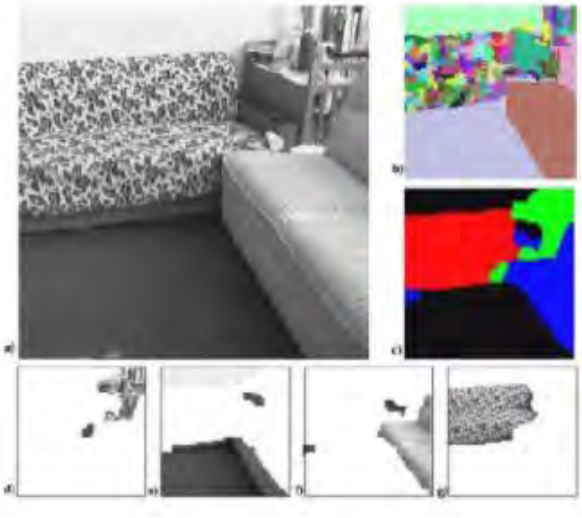


Fig 2.2.7 Segmentation based on texture

2.2.6 Object Recognition

It is the process of identifying specific objects in an image or a video. It assigns a label to an object based on its descriptors. Object recognition algorithms rely on matching, learning, or pattern recognition algorithms using appearance-based or feature based techniques [9]. It includes:

- Detection of separate objects.
- Description of their geometry and positions in 3D.
- Classification as being one of a known class.
- Identification of the particular instance.
- Understanding of spatial relationships between objects.

2.2.7 Representation & Description

Representation & Description always follows the output data from segmentation stage, which is raw pixel data.

Representation:

Boundary representation is valid when focus is on the external shape characteristics such as corners and inflections. Regional representation is appropriate when the focus is on internal properties, such as texture or skeletal shape.

Description:

It is also known as feature selection, deals with extracting attributes that results in some quantitative information of interest or are basic for differentiating one class from another[9].

2.3 Biomedical Image Processing

Over the last few decades biomedical image processing has been advanced dramatically. Biomedical images are the core of where medical science is at. It refers to the images of human body which helps in better understanding of human biological systems. It can refer to microscopic images of cell, blood vessels or can be images of complete organs, organ systems and body parts. Technological advancement in this field has been providing doctors with the tools that are needed for the most accurate diagnosis & treatment. Biomedical imaging mainly consists of X-Rays, Contrast Agents, Ultrasound, CT, and MRI etc.

- X-Rays:

It all began with the discovery of X-rays by Wilhelm Roentgen in 1895.Although diagnosis with this technology was difficult due the similarity between the densities of adjacent soft tissues within the body. It was an interesting diagnosis technique for bones or any foreign objects in the body but was not correct method for the soft tissue pathologies.

- Contrast Agents:

Between 1906 and 1912, application of pharmaceutical contrast agents which helped to visualize organs and blood vessels was a vast development in radiography. Administration of harmless contrast agents orally or via vascular injection allowed the visualization of blood vessels, digestive and gastrointestinal systems, bile ducts and gall bladder for the very first time. By 1960's, this technique was responsible for new radiological application: angiography which is used for visualizing inside or lumen of a blood vessel both arteries and veins.

- Ultrasound:

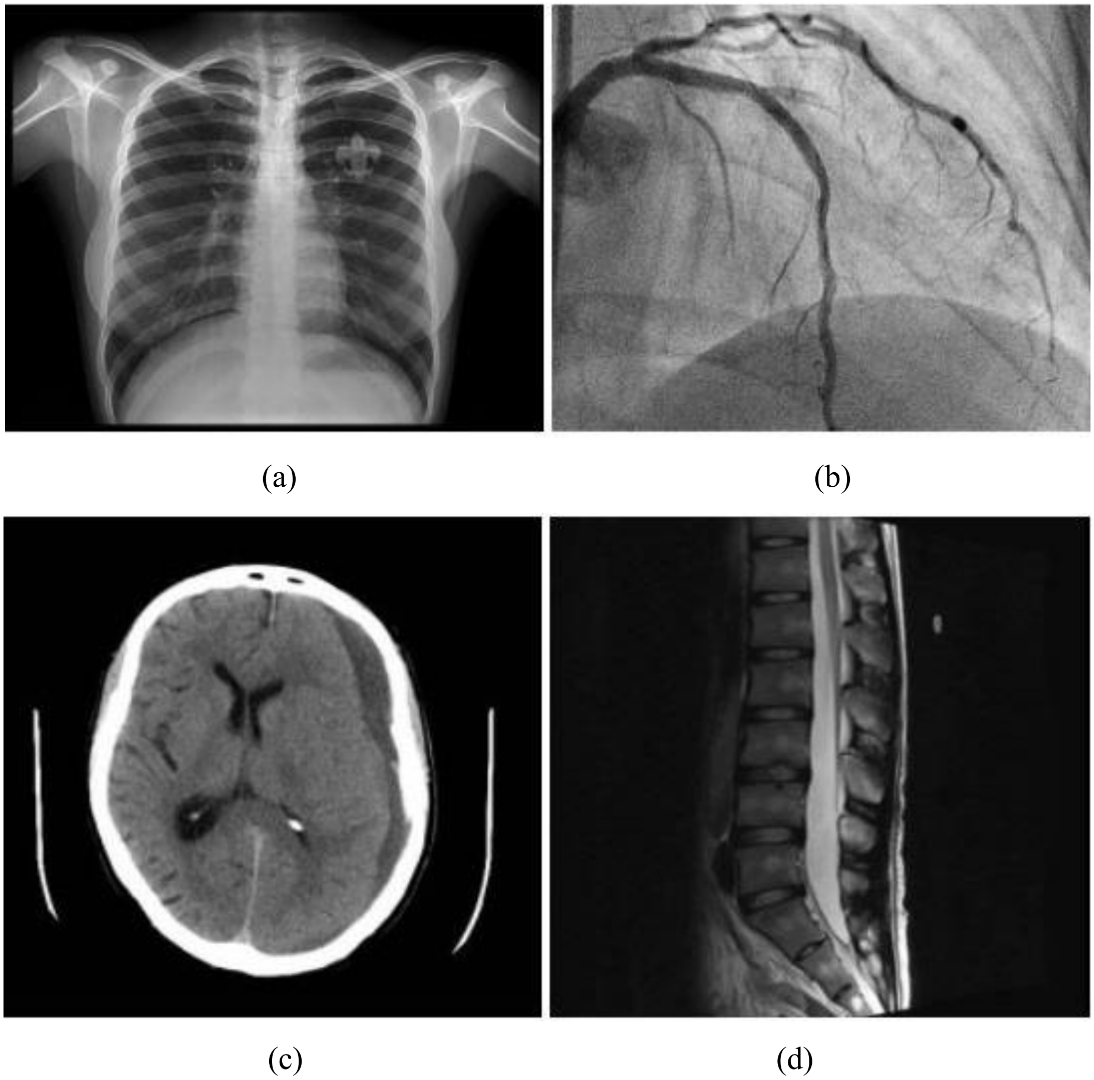
In the 1960's, the principles of Sonar (developed extensively during the Second World War) were applied to diagnostic imaging. Ultrasound is a type of sound that is above human hearing range. Ultrasound scanner transmits high frequency sound waves that penetrate into the body, bounce off the organs, and then the return sound wave vibrates transducer which turns it into electronic pulses that transforms the pulses into images. The ultrasound technology was firstclinicallyused inthe1970s.

- Computed Tomography(CT) :

Godfrey Hounsfield and Allan Cormack were awarded Nobel Prize in Medicine for inventing computed tomography (CT) in 1972. CT scan refers to the computer processed X-rays (taken from different angles) which can produce tomographic images of specific areas of a scanned object. It provides better insight of body without cutting it.

- Magnetic Resonance Imaging(MRI):

Magnetic Resonance Imaging was first used in 1970's and continued evolving throughout the years. Over 1980's it has become common in medical imaging. Briefly, it is a tomographic imaging technique that produces images of internal physical and chemical characteristics of an object from externally measured nuclear magnetic resonance (NMR) signals. MR imaging is based on the well-known NMR phenomenon[14].



(a) (b)

(c) (d)

Fig. 2.3.1 (a) Chest X-ray (b) Angiogram (c) CT scan of intracranial hemorrhage (d) MRI of Spine

2.3.1Analysis of the microscopy image collected in Biopsy

In general, tissue and cell samples are examined by pathologists, doctors with special training in diagnosing diseases by lab tests. Sometimes, other doctors will also examine specimens or tissues of organs related to their area of expertise.

The examiners will mostly consider the following while conducting a biopsy of a tissue or cell:

- <u>Size and shape of the cells:</u>

The overall size and shape of the cancer cells are usually abnormal. They can be either smaller or larger than the normal cells. Normal cells often have defined shapes that help them do their tasks. Cancer cells usually do not function in a useful way and their shapes are often distorted. Unlike normal cells that tend to have the similar size and shape, cancer cells vary greatly in shapes and sizes.

- Size and shape of the cell's nucleus

  The nucleus is the control center of the cell that contains the cell's DNA. The shape and size of the nucleus of a cancer cell is often abnormal. Typically, the nucleus of a cancer cell is enlarged and darker than that of a normal cell nucleus and its size may vary greatly.

- Arrangement of the cells

  The arrangement of normal cells reflect the function of each tissue. For example, cells can form glands that make substances that are taken to other parts of the tissue.

  Cancerous cells grow into other tissues. Normal cells stay where they belong within a tissue. The ability of cancer cells to invade reflect the fact that their growth and movement is not very coordinated with their neighboring cells. This capability to invade is how cancer cells spread to and damages nearby tissues. And, unlike normal cells, cancer cells can metastasize (spread through lympor blood vessels) to other parts of the body.

- The type of cancer

  There are several basic kinds of cancers, which doctors can further classify into hundreds or even thousands of types, based on how they look under a microscope. Cancers are classed according to which type of normal cells and tissues they look like most. For example, cancerous tissue that look like glandular tissues are called adenocarcinomas. Other cancers that resemble certain immune system cells are called lymphomas, and those that look like bone or fat tissue are osteosarcomas and liposarcomas, respectively.

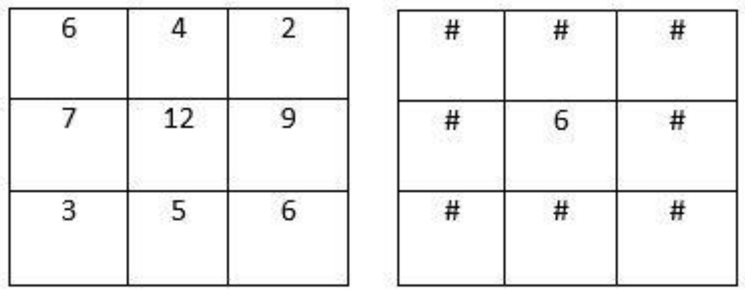
- Grading the cancer

While identifying the cell type or tissue a cancer looks like, doctors also decide how closely they look like the normal cells or tissues. This is the grade of the cancer. Cancers that look more like normal tissues are called low grade, and those that don’t look much like normal tissues are high grade. A high-grade cancer tends to grow and spread faster than a low-grade cancer.

2.4 Image Processing Techniques Related to Cancer Cell Detection

2.4.1 Mean Filter

Mean filter is used for smoothing & reducing noises from an image. The working procedure of mean filter is very simple. It replaces the pixel values in an image with the average of its neighbor pixels. This filter cancels out the pixel with unrepresentative surroundings. Mean filter works based on a kernel representing size and shape of the neighborhood when determines the mean [15, 16]. Usually a square kernel is taken but it can be of any shape.
An example of mean filter is shown below.



6	4	2
7	12	9
3	5	6

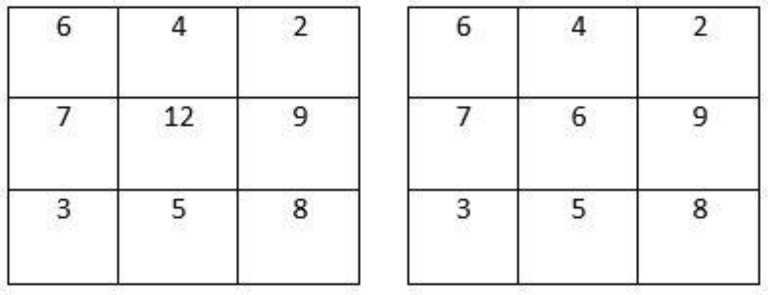
#	#	#
#	6	#
#	#	#

(a) (b)

Fig. 2.4.1 (a) Matrix of pixel (b) Mean Filtered matrix of pixel

2.4.2 Median Filter

Just like mean filter, median filter is also used to reduce the amount of noise in an image. But it has two major advantages over mean filter. Firstly is gives more vigorous average of the pixels than mean filter. Secondly, it does not create unrealistic pixel value while filtering. Thus gives us sharper edges than mean filtered image [15]. Median filter also follows a simple working procedure. For every pixel in an image, its neighbor pixels are considered. First the neighboring pixels are sorted out and then the pixel is replaced by the median value of all the sorted pixels and if the neighboring pixels are even in number, then the average value of the middle two pixel value is used for the replacement [16]. An illustration of the calculation of median filtering is given below.



6	4	2
7	12	9
3	5	8

(a)

6	4	2
7	6	9
3	5	8

(b)

Fig. 2.4.2 (a) Matrix of pixel (b) Median Filtered matrix of pixels

Here, the center pixel value is 12. If the neighboring pixels are being sorted, then the sorted list will be as: 2, 3, 4, 5, 6, 7, 8, 9, and 12. The median value of the list is 6. Hence the center value (previously 12) will be replaced by 6.

2.4.3 Gaussian Smoothing Filter

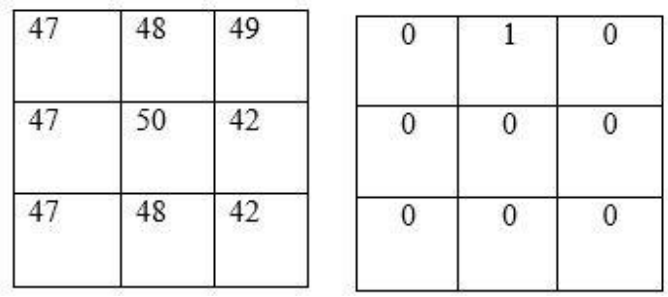
The Gaussian smoothing operator is a 2-D convolution operator that uses a Gaussian function for calculating the transformation in each pixel. It is widely used to 'blur' images and also for removing noises and details from an image. The equations of Gaussian function in both 1D and 2D are given below:

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

$$G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

Here *x* is the distance from the origin in the horizontal axis, *y* is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

Gaussian smoothing uses 2-D distribution function and when calculating new values of a given pixel, convolution operator id being used. Convolution takes values of the neighboring pixels into account. The main element of a convolution is kernel which is a matrix of arbitrary size mostly a square matrix. When calculating the new value of the selected pixel, the convolution kernel is applied to it by its center pixel. Neighboring pixels are covered with the same kernel. Next, the sum of the product of the pixels is calculated in the image. The resulting sum is the new value of the selected pixel. Now, if the convolution is applied to each pixel in the image, we get a certain effect, which depends on the chosen convolution kernel. An example of the calculation of Gaussian Smoothing is given below.



47	48	49
47	50	42
47	48	42

0	1	0
0	0	0
0	0	0

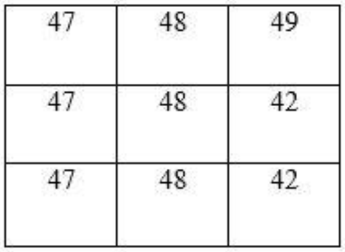
(a) (b)

Fig. 2.4.3 (a) Matrix of the image (b) Convolution Kernel

The result is calculated in following way:

Result= (47*0) + (48*1) + (49*0) + (47*0) + (50*0) + (42*0) + (47*0) + (48*0) + (42*0) = 48.

The result of applying kernel to a pixel with center value of 50 will be:



47	48	49
47	48	42
47	48	42

Fig. 2.4.4 Gaussian smoothing filtered matrix of the image

2.4.4 Distance Transform Algorithm

The distance transformed algorithm is used for transforming a binary image to a graylevel image where graylevel intensities within the object or interior point are changed for showing the distance to the closest boundary from each pixel. There are different types of distance transform such as Euclidean distance, City block transform distance, and Chessboard distance. These distance metrics are used to determine distance between pixels. The formula of distance transform is given below:

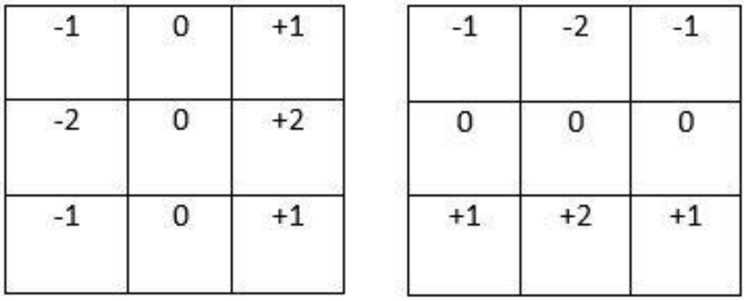
Set of points is P, and measure of distance:

$$DT\ (P)[x] = min_{y \in P}\ dist(x, y) \tag{3}$$

Here, for each location 'x' distance to nearest point is 'y' in *P*.

2.4.5 Sobel Edge Detection

The Sobel Detection is used to detect the edges in an input grayscale image. The Sobel operator performs a 2-D spatial gradient measurement on an image and creates an edge emphasized image. The sobel operator consists of two 3x3 convolution kernels where one kernel is rotated by 90°.



-1	0	+1
-2	0	+2
-1	0	+1

-1	-2	-1
0	0	0
+1	+2	+1

G_x G_y

Fig. 2.4.5 Sobel Convolution Kernel

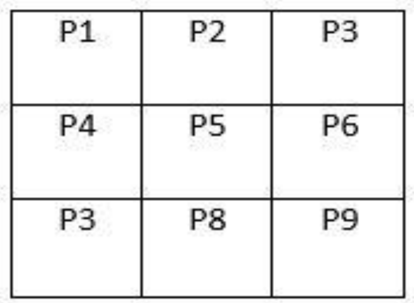
The kernels are designed in a way so that they respond to vertical and horizontal edges. The gradient components (called G_x andG_y) can be measured individually by applying the kernels individually to the image [17]. Absolute magnitude of the gradient can be found from these components (G_x andG_y). The gradient magnitude is:

$$|G| = \sqrt{Gx^2 - Gy^2} \tag{4}$$

Gradients direction can be calculated as well from this information.

$$\theta = \tan^{-1}\frac{G_y}{G_x} \tag{5}$$

The user often visualizes this absolute magnitude as the only output. The two components of the gradient are calculated and added in a single pass over the input image using the pseudo-convolution operator [18].



P1	P2	P3
P4	P5	P6
P3	P8	P9

Fig.2.4.6Pseudo-convolution kernel

Using this kernel, approximate gradient magnitude is,

$$|G| = |\,(P1 + 2 \times P2 + P3) - (P7 + 2 \times P8 + P9)|+|\,(P3 + 2 \times P6 + P9) - (P1 + 2 \times P4 + P7)\,| \tag{6}$$

2.4.6 Canny Edge Detection

It is a popular algorithm for detecting edge developed by John F. Canny. It is a multi-step algorithm that reduces noise and detects edges at the same time. First the noises of the image are removed by a 5x5 Gaussian filter. The next step is to find the intensity gradient of the image. To find that, the noise filtered image is again filtered with Sobel kernel in both horizontal and vertical direction to get gradient components (called G_x andG_y). From these, edge gradient and direction for each pixel can be found using the following equation:

$$|G| = \sqrt{Gx^2 - Gy^2} \tag{7}$$

$$\theta = \tan^{-1}\frac{G_y}{G_x} \tag{8}$$

After that, unwanted pixels which may not be a part of the edge are removed by a full scan of the image. For doing this, pixel is checked at every pixel if it is a local maximum in its neighborhood in the direction of gradient and we get a binary image with thin edges. The next step is Hysteresis Thresholding. This step determines whether edges are really edges or not. For this, threshold values, maxVal and minVal are needed. Edges having intensity gradient more than maxVal are detected as edge and those below minVal are sure to be non-edge. Those living between these two thresholds are classified edges or non-edges based on their connectivity. If they are connected with the edge pixel, they are treated as part of the edges and if they are not connected, they are non-edges.

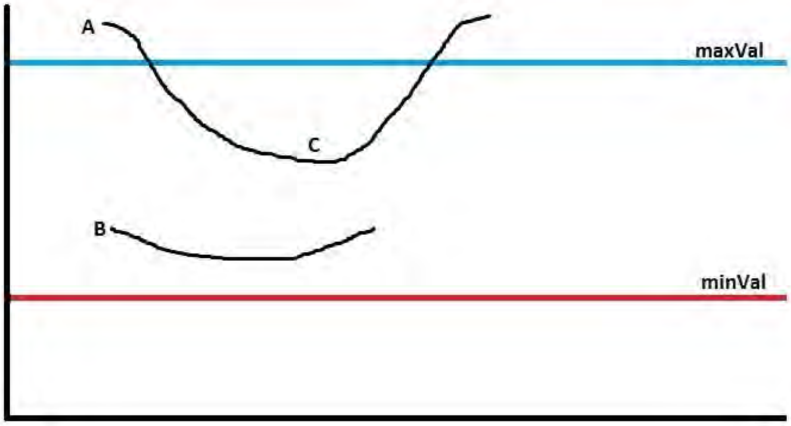
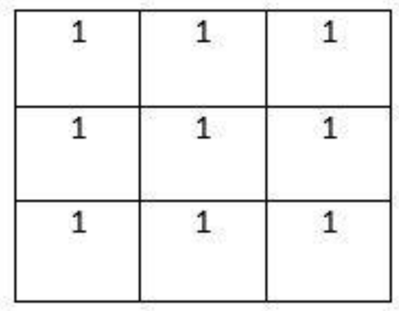


Fig. 2.4.7 Hysteresis Thresholding

The point A here is above maxVal. So it is considered as edge. Point C is also considered to be a part of the edge as it is connected to the edge pixel. On the other hand, point B is considered as non-edge as it is lie between maxVal and minVal and it is not connected to any edge pixel.

2.4.7 Erosion

Erosion and dilation are two fundamental operations to morphological processing. Erosion is usually applied to binary and sometimes in grayscale images. It is used on a binary image in order to remove the boundaries of the regions of the foreground pixels. Thus erosion shrinks the objects in a binary image. The erosion operator takes two sets as input, for example A and B where A is the set of coordinates corresponding to the binary image and B is assumed to be the set of structuring element. Then the erosion A by B is the set of all points (for instance call it ‘*x*’) such that the set of points ‘*x*’ is contained in A. This structuring element determines the effects of the erosion on the image. For example of binary erosion operation, assume a structuring elements as a 3x3 square with origin at (0, 0) shown in the figure below where foreground pixels are 1’s and background pixels are 0’s.



1	1	1
1	1	1
1	1	1

Fig. 2.4.8 Structuring Element (3x3)

Set of coordinate points={(-1,-1),(0,-1),(1,-1),(-1,1),(0,0),(1,0),(-1,1),(0,1),(1,1)} For each foreground pixel, the structuring element placed on top of the input image in order to coincide the origin of structuring element with input pixel coordinates. For all the pixels in the structuring element, if the corresponding pixel in the image underneath is a foreground pixel, then the input pixel remains untouched. If any of the corresponding pixels in the image are background then the input pixel is also set to background value. For the above example, the effect of this operation is to remove any foreground pixel that is not completely surrounded by other white pixels and such pixels will lie at the edges of the white pixels so that the foreground region thins or shrinks.

2.4.8 Marker Controlled Watershed

The segmentation is a very important part of image processing. There are two major ways of segmentation of an image. One is the frontier approach and the other being region approach. Watershed segmentation combines both the approaches and thus gives a fast result detecting both edges and regions [19, 20]. But main drawback with the watershed segmentation is that it leads to over-segmentation of regions and most of the time, the segmented image seems very noisy. Hence, it becomes useless sometimes. To overcome this problem, markers are used for segmentation to reduce the number of regional minima. A marker is a connected component belonging to an image. The markers which are connected components do possess the same intensity values and are treated as regional minima. Markers can be of foreground or background depending on its location from region of interest. The use of markers gives a prior knowledge about segmentation; this knowledge often consists in the number and the positions of the regions through the definition of some markers, and thus simplifies the problem of over-segmentation [21, 22].

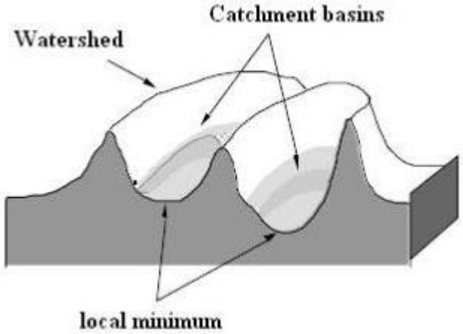


Fig. 2.4.9 Watershed Transform

2.4.9 OTSU's Threshold Algorithm

OTSU's algorithm is an image binarization algorithm named after its inventor Nobuyuki Otsu. This algorithm assumes that an image has two types of pixels (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread is minimal, or equivalent, so that their inter-class variance is maximal [23]. In this method, a threshold value is searched that minimizes intra-class variance, defined as a weighted sum of variances of the two classes:

$$\sigma_{\omega}{}^{2}\,(t)\ =\ \omega_0(t)\,\sigma_0{}^{2}(t)\ +\omega1(t)\,\omega_1{}^{2}(t) \tag{9}$$

Weights ω_0 and ω_1 are the probabilities of the two classes separated by a threshold ***t*** and $\boldsymbol{\sigma}_0{}^2$ and $\boldsymbol{\sigma}_1{}^2$ are variances of these two classes. The class probability $\omega_{0,1}(t)$ is computed from the ***L*** bins of the histogram:

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i) \tag{10}$$

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i) \tag{11}$$

Otsu shows minimizing the intra-class variance and maximizing inter-class variance is equivalent.

$$\begin{aligned} \sigma_b{}^2(t) &= \sigma^2 - \sigma_\omega{}^2(t) \\ &= \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \\ &= \omega_0(t)\omega_1(t)[\,\mu_0(t) - \mu_1(t)]^2 \end{aligned} \tag{12}$$

This is expressed in terms of class probabilities ***w*** and class means ***μ,*** while the class mean $\boldsymbol{\mu}_{\mathit{10,1,T}}\boldsymbol{(t)}$ is:

$$\mu_0(t) = \sum_{i=0}^{t-1} i\,\frac{p(i)}{\omega_0} \tag{13}$$

$$\mu_1(t) = \sum_{i=t}^{L-1} i\,\frac{p(i)}{\omega_1} \tag{14}$$

$$\mu_T = \sum_{i=0}^{L-1} ip(i) \tag{15}$$

In Matlab, we used the built-in function *graythresh()* which is implemented with Otsu's algorithm.

2.4.10 Connected Components Labeling

Connected components labeling detects connected regions by scanning an image based on pixel connectivity. Pixels in connected components have some pixel intensity values in common and they are someway connected with one another. After determining the connected regions, all the pixels are labeled with a color. Connected components labeling scans an image from top to bottom and from left to right, pixel by pixel for identifying connected pixel group for example regions of adjacent pixels which share the same set of intensity values. The connected components labeling operator scans the image by moving along a row up to a point *p* (where *p* denotes the pixel to be labeled at any stage in the scanning process) for which *V= {1}*. When this is true, it examines the four neighbors of *p* which have already been encountered in the scan.

Based on this information, the labeling of *p* occurs as follows:

- If all four neighbors are 0, assign a new label to *p*, else
- if only one neighbor has V={1}, assign its label to *p*, else
- If more than one of the neighbors have *V= {1}*, assign one of the labels to *p* and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class. As a final step, a second scan is made through the image, during which each label is replaced by the label assigned to its equivalence classes. For display, the labels can be of different colors.

2.4.11 Douglas-Peucker Algorithm

This is a line simplification algorithm and its purpose is to find a curve with fewer points from a given curve composed of line segments. The curve with fewer points consists a subset of points of original curve. The algorithm recursively divides the line. At first, it is given all the points between the first and last point. It keeps the first and last point by marking them automatically. After that, it finds the point that is furthest from the line segment with the first and last points as end points; this point is the furthest on the curve from the approximating line segment between the end points. If the point is closer than *ε (Epsilon)* to the line segment, then any points that are not currently marked are discarded without the simplified curve being worse than *ε (Epsilon)* [24].

If the point furthest from the line segment is greater than *ε (Epsilon)* from the approximation then that point must be kept. The algorithm recursively calls itself with the first point and the furthest point and then with the furthest point and the last point, which includes the furthest point being marked as kept.

After the completion of recursion, a new curve is generated with those points that have been marked for keeping.

Chapter 3
Proposed System

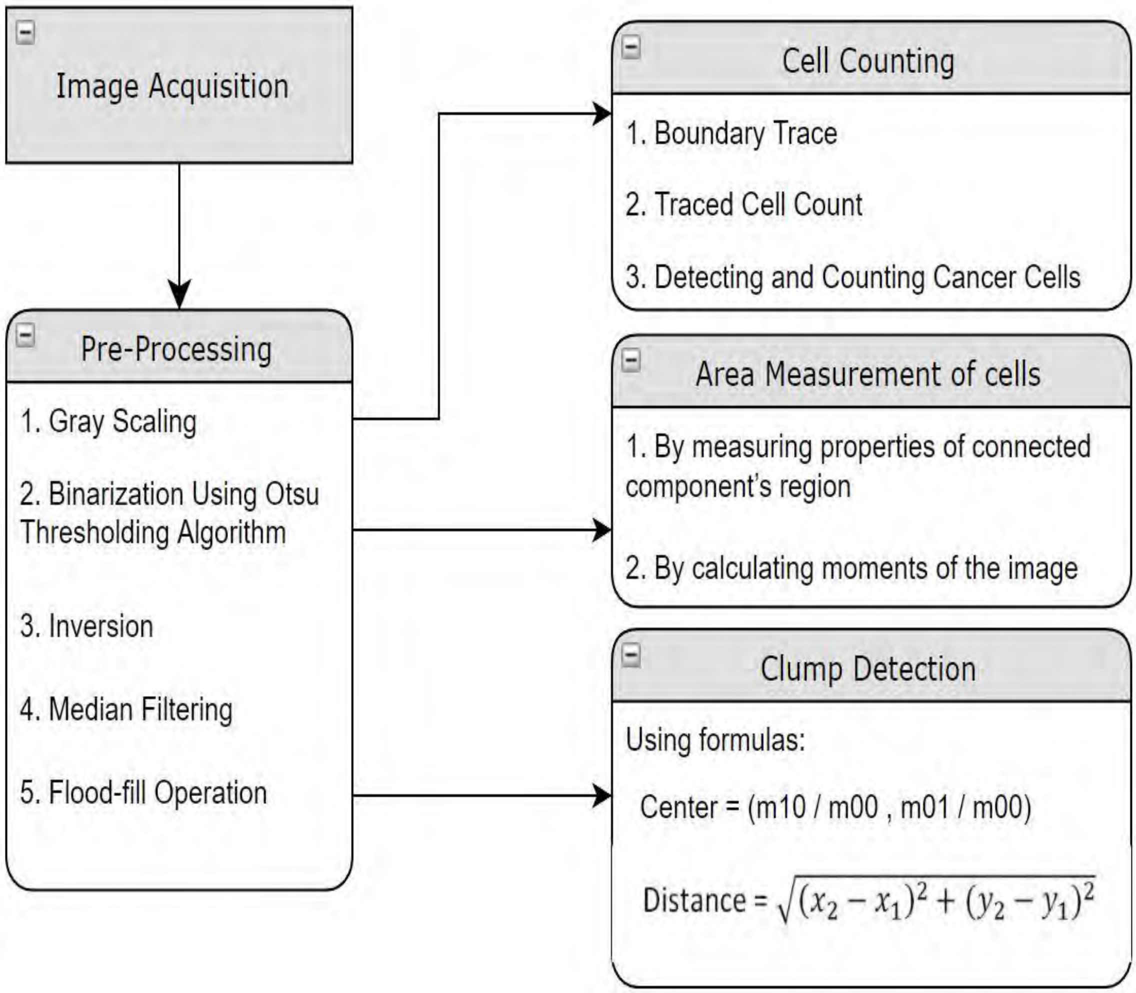


Image Acquisition
Pre-Processing
1. Gray Scaling
2. Binarization Using Otsu Thresholding Algorithm
3. Inversion
4. Median Filtering
5. Flood-fill Operation
Cell Counting
1. Boundary Trace
2. Traced Cell Count
3. Detecting and Counting Cancer Cells
Area Measurement of cells
1. By measuring properties of connected component's region
2. By calculating moments of the image
Clump Detection
Using formulas:
Center = (m10 / m00 , m01 / m00)
Distance = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Fig. 3.1 Flowchart of the proposed system

3.1 Pre-Processing

Before detecting cancer via image processing techniques, the acquired image needs to undergo several filtering to make the image easier to process and generate more accurate results.

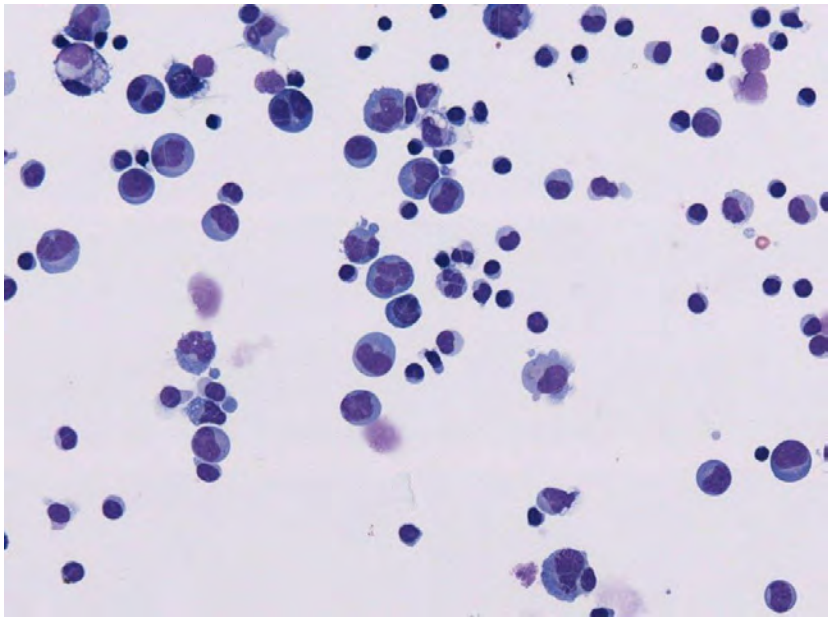


Fig. 3.1.1 Biopsy image collected from the microscope
Via a macro camera

3.1.1 Gray Scaling

The image is converted to grayscale to reduce color complexity and noise. It is done by rejecting hueand saturation information and using the following equation.

$$(0.2989 * R) + (0.5870 * G) + (0.1140 * B) \qquad (16)$$

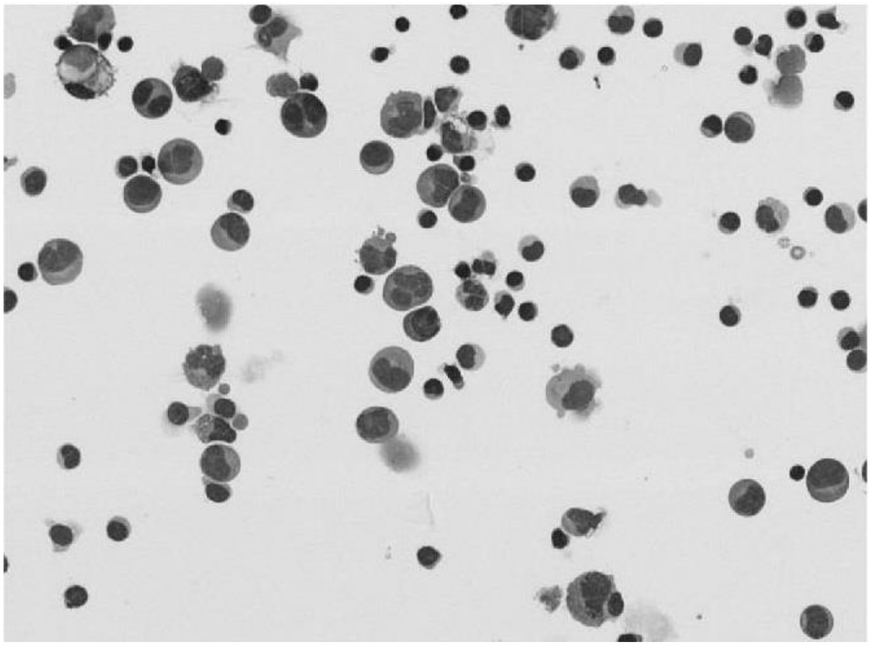


Fig. 3.1.2 Image after Gray Scale

3.1.2 Binarization Using Otsu thresholding algorithm

OTSU's algorithm is an image binarization algorithm named after its author Nobuyuki Otsu. This algorithm assumes that an image has two types of pixels (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread is minimal, or equivalent, so that their inter-class variance is maximal [22]. In this method, a threshold value is searched that minimizes intra-class variance, defined as a weighted sum of variances of the two classes:

$$\sigma_{\omega}{}^{2}\ (t)\ =\ \omega_0(t)\,\sigma_0{}^{2}(t)\ +\omega 1(t)\,\omega_1{}^{2}(t) \qquad (17)$$

Weights $\omega_{\boldsymbol{0}}$ and $\omega_{\boldsymbol{1}}$ are the probabilities of the two classes separated by a threshold $\boldsymbol{t}$ and $\boldsymbol{\sigma}_0{}^2$ and $\boldsymbol{\sigma}_1{}^2$ are variances of these two classes. The class probability $\omega_{0,1}(t)$ is computed from the $\boldsymbol{L}$ bins of the histogram:

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i) \qquad (18)$$

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i) \qquad (19)$$

Otsu shows minimizing the intra-class variance and maximizing inter-class variance is equivalent.

$$\begin{aligned}\sigma_b{}^2(t) &= \sigma^2 - \sigma_\omega{}^2(t)\\ &= \omega_0(\mu_0-\mu_T)^2 + \omega_1(\mu_1-\mu_T)^2\\ &= \omega_0(t)\omega_1(t)[\,\mu_0(t)-\mu_1(t)]^2 \qquad (20)\end{aligned}$$

This is expressed in terms of class probabilities $\boldsymbol{w}$ and class means $\boldsymbol{\mu}$, while the class mean $\boldsymbol{\mu_{10,1,T}(t)}$ is:

$$\mu_0(t) = \sum_{i=0}^{t-1} i\,\frac{p(i)}{\omega_0} \qquad (21)$$

$$\mu_1(t) = \sum_{i=t}^{L-1} i\,\frac{p(i)}{\omega_1} \qquad (22)$$

In Matlab, we used the built-in function *graythresh()* which is implemented with Otsu's algorithm. The grayscale image is converted to binary image with level value 0.55.

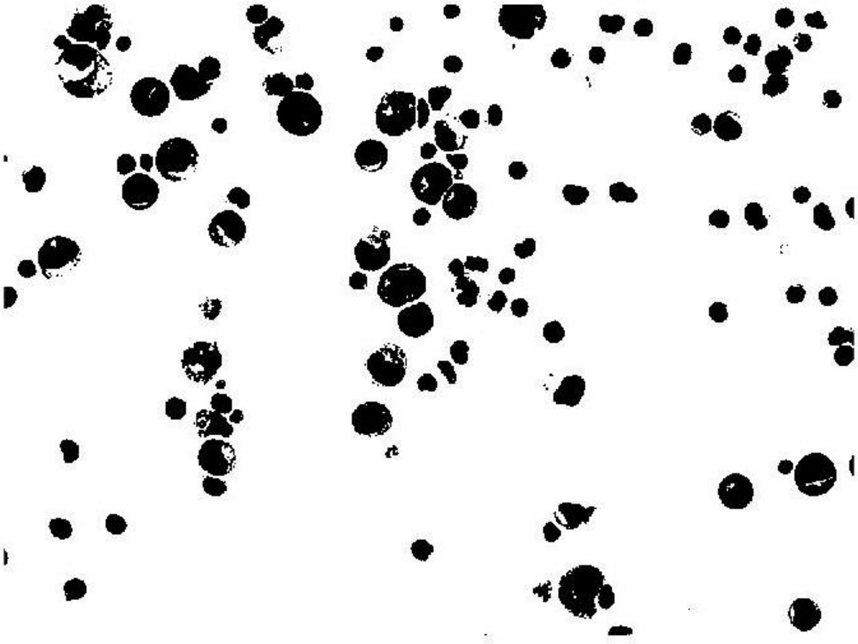
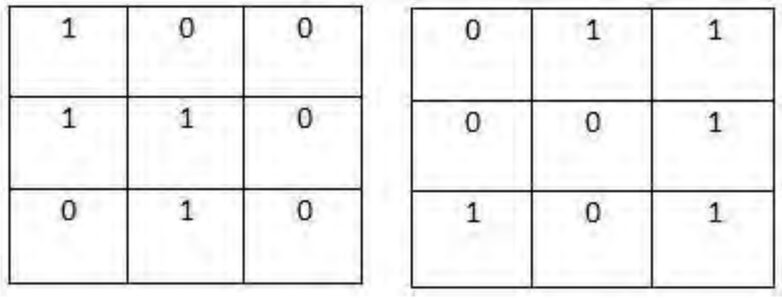


Fig. 3.1.3 Binary Image converted from Gray Scale

3.1.3 Inversion

The binary value of each pixel is inverted using the built-in inversion command in matlab.



1	0	0
1	1	0
0	1	0

(a)

0	1	1
0	0	1
1	0	1

(b)

Fig. 3.1.4 (a) Pixel of Image (b) Pixel of Inverted Image

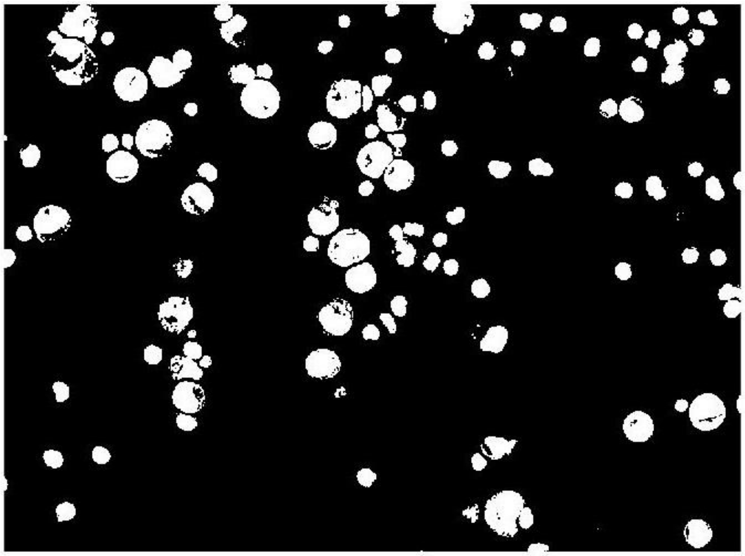
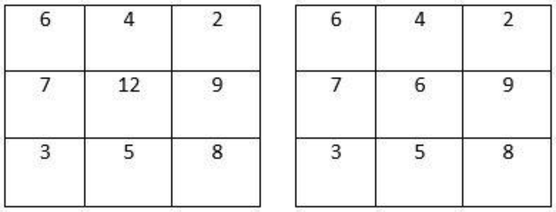


Fig.3.1.5 Inverted Image converted from Binary Image

3.1.4 Median Filtering

The purpose of this filtering is to reduce unwanted noise. After performing Median filtering for a suitable number of times, unwanted noise is significantly reduced. It gives vigorous average of the pixels. It does not create unrealistic pixel value while filtering. Thus gives us sharper edges [15]. Median filter also follows a simple working procedure. For every pixel in an image, its neighboring pixels are considered. First the neighboring pixels are sorted out and then the pixels are replaced by the median value of all the sorted pixels and if the neighboring pixels are even in number, then the average value of the middle two pixel value is used for the replacement [16]. An illustration of the calculation of median filtering is given below.



6	4	2
7	12	9
3	5	8

(a)

6	4	2
7	6	9
3	5	8

(b)

Fig. 3.1.6 (a) Matrix of pixel (b) Median Filtered matrix of pixels

Here, the center pixel value is 12. If the neighboring pixels are being sorted, then the sorted list will be as: 2, 3, 4, 5, 6, 7, 8, 9, and 12. The median value of the list is 6. Hence the center value (previously 12) will be replaced by 6.

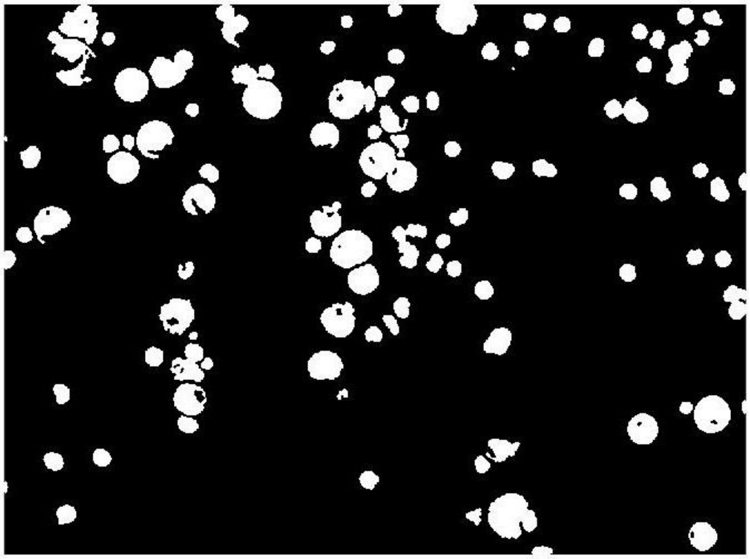


Fig. 3.1.7 Image after Median Filtering

3.1.5 Flood-fill operation

The objects (in this case, the cells) in the binary image needs to be a solid otherwise multiple cells will be counted instead of one. These can be done by filling the background regions using morphological reconstruction which recover the minima that are not connected with in an object boundary .This flood-fill operation area applied the binary image where it changes the connected background pixels to foreground. Mainly this operation will be occurred till it reaches the boundaries.

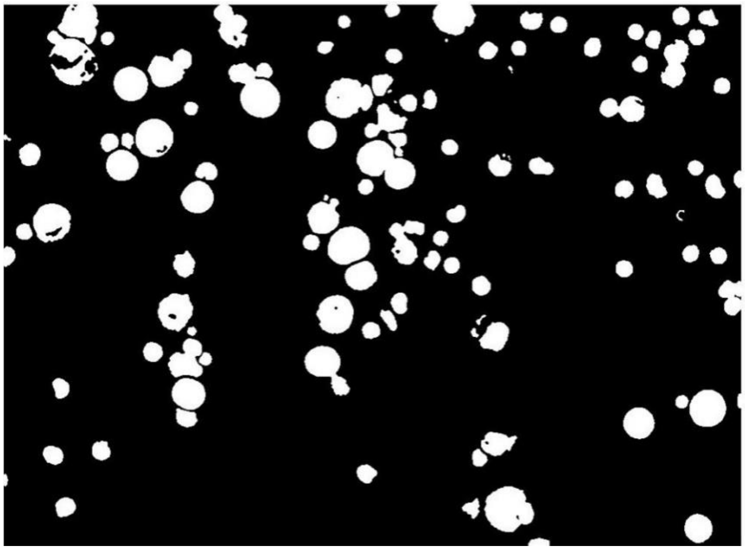


Fig. 3.1.8 Image after Flood Fill Operation

3.2 Cell Counting

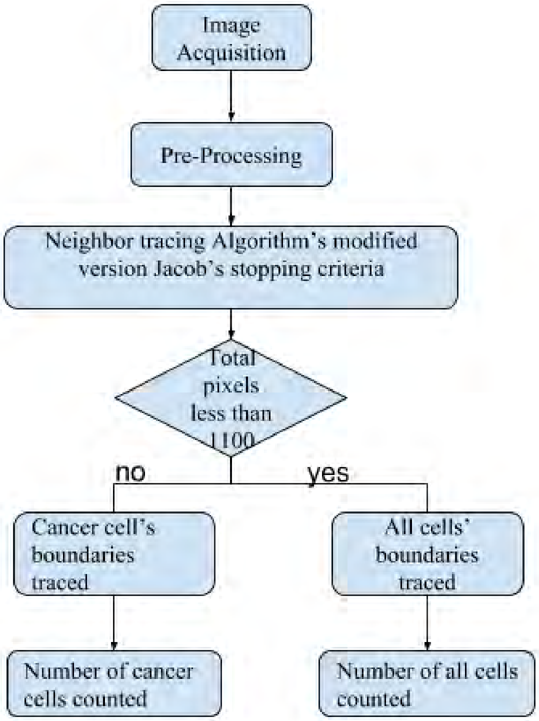


Image Acquisition
Pre-Processing
Neighbor tracing Algorithm's modified version Jacob's stopping criteria
Total pixels less than 1100
no
yes
Cancer cell's boundaries traced
All cells' boundaries traced
Number of cancer cells counted
Number of all cells counted

Fig. 3.2.1 Flowchart of cell counting

3.2.1 Boundary trace

To trace the boundaries or the outline of the cells, we applied Moore-Neighbor tracing Algorithm's modified version Jacob's stopping criteria which finds the Moore neighborhood of the current boundary pixel in a certain direction until it finds a black pixel. It then declares that pixel as the present boundary pixel and does the same process again.

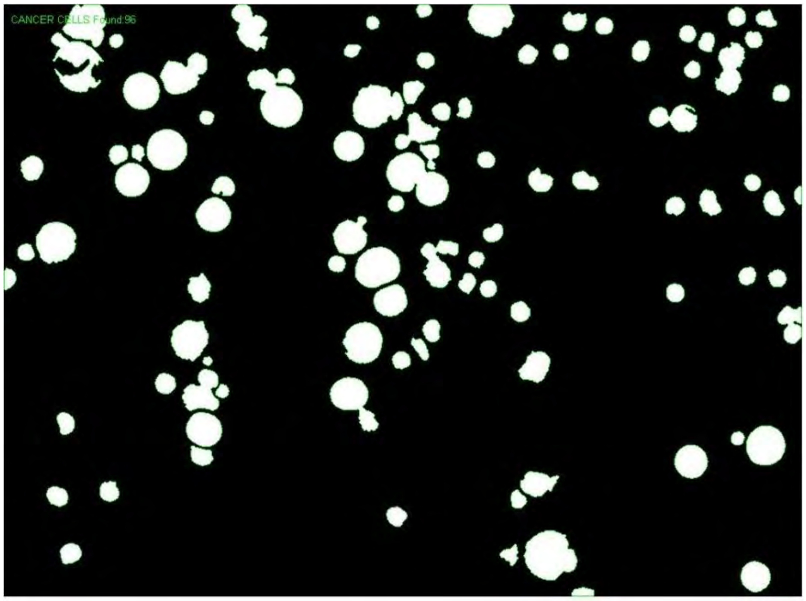


Fig. 3.2.2 Marked and counted all cells

3.2.2 Traced cell counting

Now, length of the returning value from the traced boundary is the counted cells and displayed by plotting based on row and column coordinates gained by applying the algorithm with fix line thickness and color.

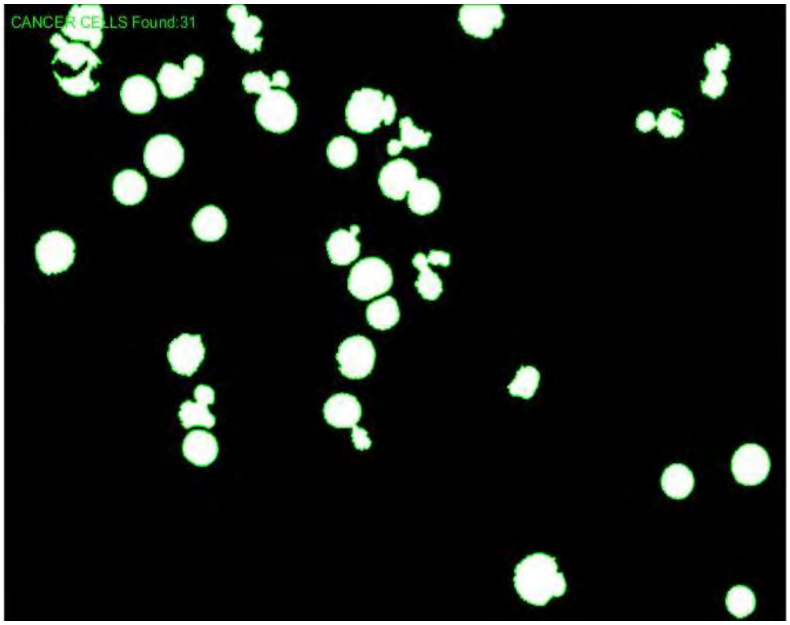


Fig. 3.2.3 Marked and counted cancer cells in binary form the cell

3.2.3 Detecting and counting cancer cells

Using binary area opening method, cells that have a size of fewer than 1100 pixels (for this particular image) are removed and the remaining cells are counted as the cancer cells.

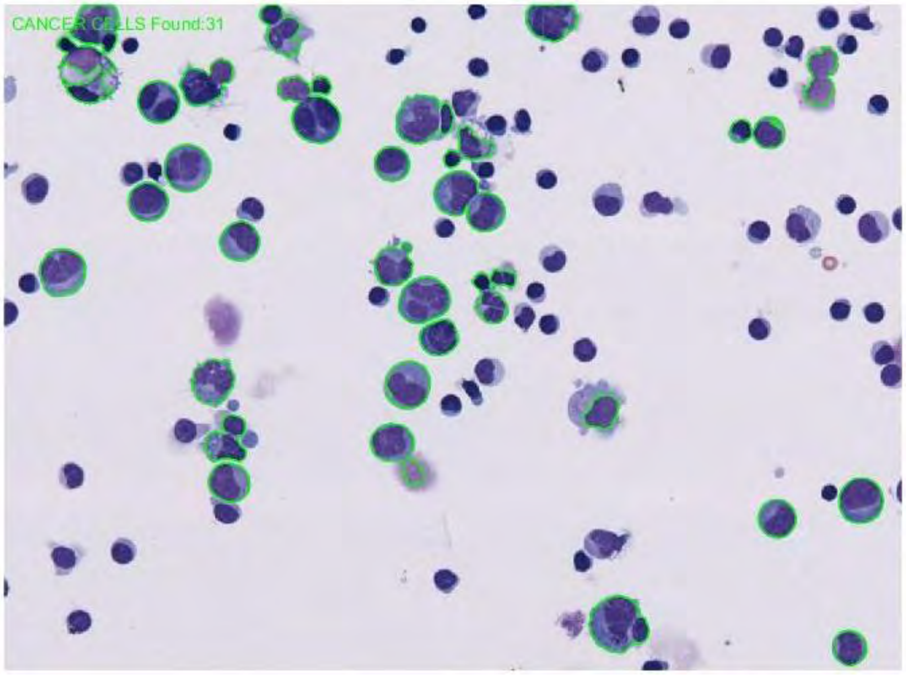
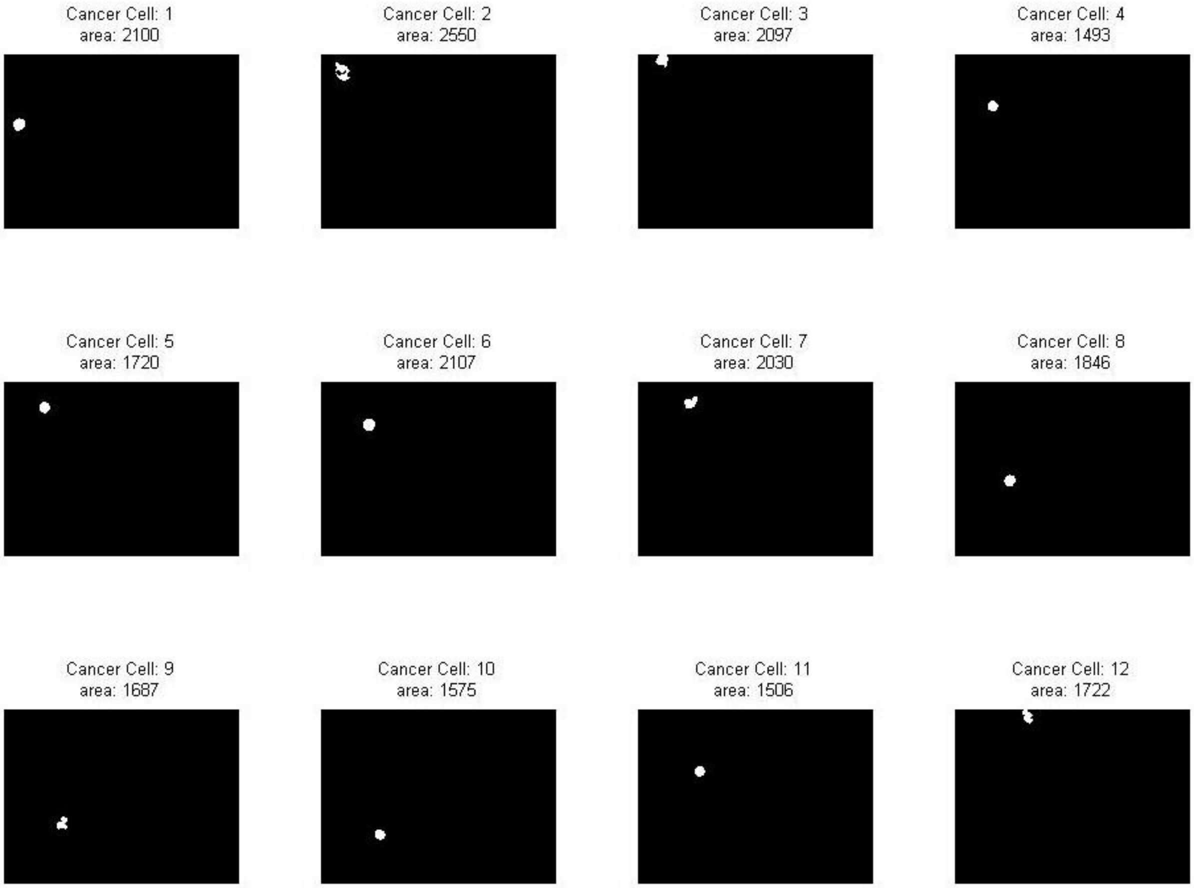


Fig. 3.2.4 Marked and counted cancer cells

3.3Measuring Cancer and Normal Cell Area and Center

3.3.1 First Approach for Area Measuring

After counting the cells, we follow same pre-processing procedure .Further, the method of connecting components are used to find connected regions and how many pixels are connected. As it is 2D image and need to find pixels that are directly connected or touches one of their edges, we are using 4-connected neighborhood.



Cancer Cell: 1
area: 2100

Cancer Cell: 2
area: 2550

Cancer Cell: 3
area: 2097

Cancer Cell: 4
area: 1493

Cancer Cell: 5
area: 1720

Cancer Cell: 6
area: 2107

Cancer Cell: 7
area: 2030

Cancer Cell: 8
area: 1846

Cancer Cell: 9
area: 1687

Cancer Cell: 10
area: 1575

Cancer Cell: 11
area: 1506

Cancer Cell: 12
area: 1722

Fig. 3.3.1 Cancer cells (1-12) cropped and labeled with area

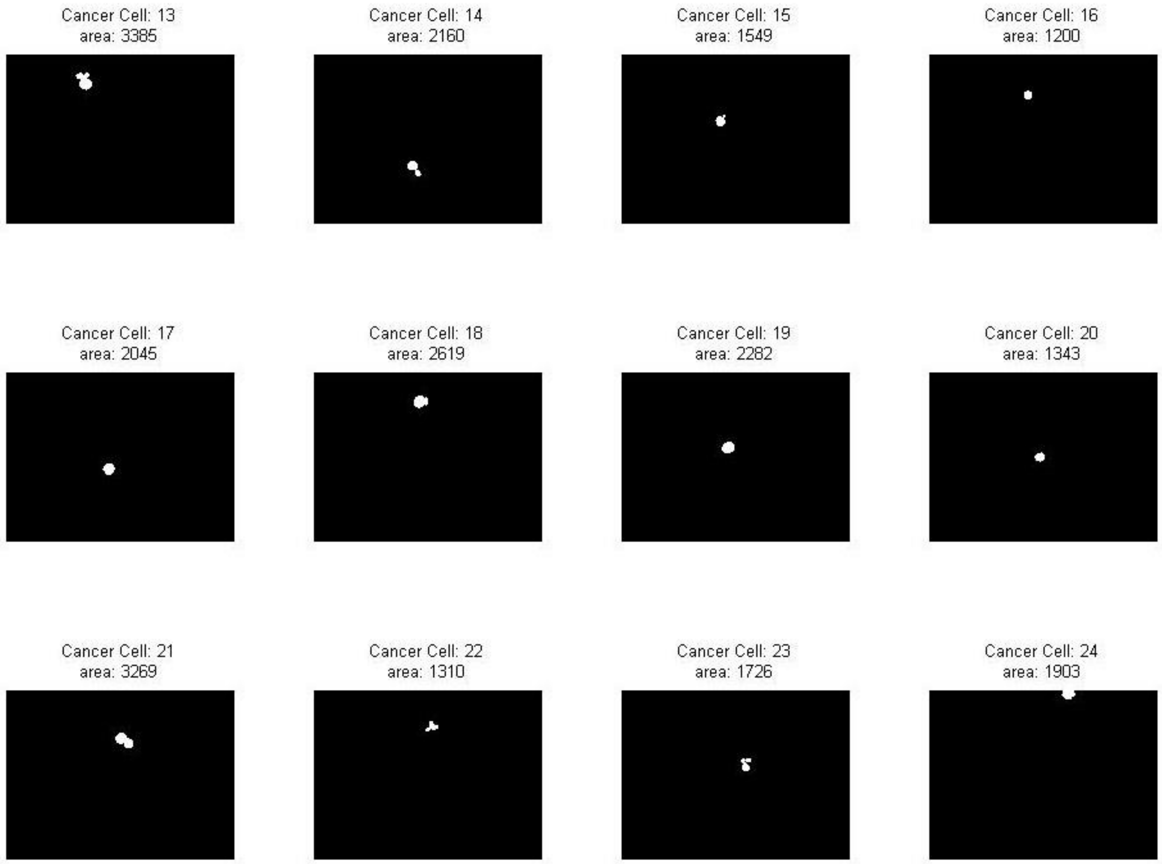


Fig. 3.3.2 Cancer cells (12-24) cropped and labeled with area

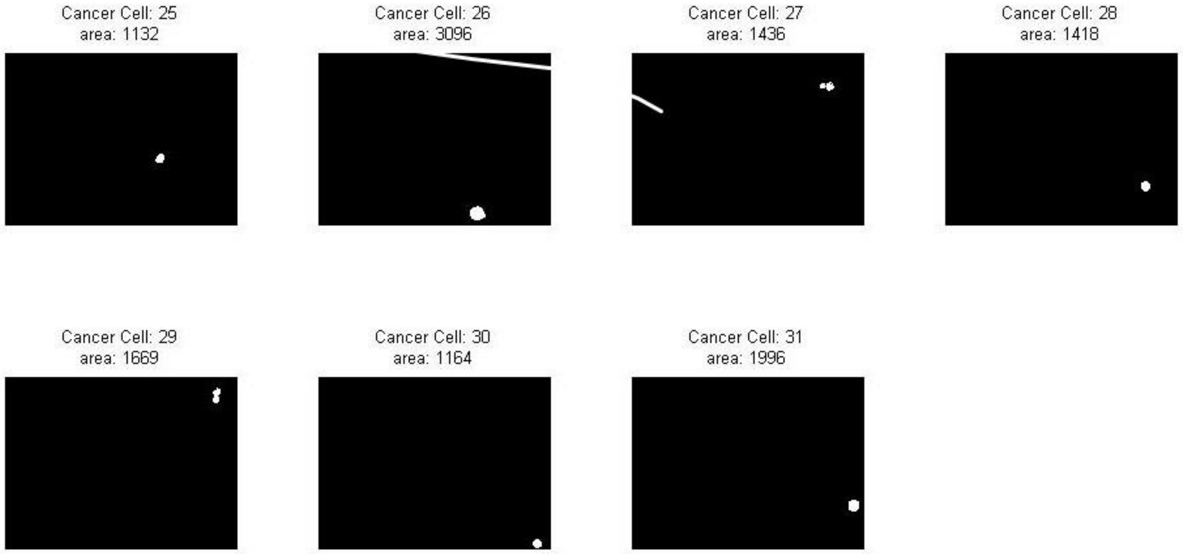


Fig. 3.3.3 Cancer cells (25-31) cropped and labeled with area

Now, in order to find the area, we measure properties of the regions using matlab'sregionprop function. This Returns a scalar that specifies the actual number of pixels in the region. The Contiguous regions are the cells. The total number of non-zero white pixels in connected region is the area of cells. Then, by our examination on this particular image, we split cancerous and non-cancerous cells based on area or total pixels in a region under 1100 or above. If area is under 1100, then is non-cancerous cell, else it if cancer cell.

3.3.2 Second Approach for Area Measuring

Area of a cell is the total number of non-zero pixels in a bounded area. To reduce complexity processing the image, we convert it to grayscale. Then, by applying Gaussian smoothing we reduce noise and details. Now, before finding the contour, we binarize the image using ostu threshing algorithm. As we are using 2-D image, the image may content many lower level cells and other parts that can be counted as noise. Mean filtering can be used to reduce this noises.We perform Contour approximation method to bounded cells. Contour approximate method which is the implementation of Douglas-PeuckerAlgorithm, stores all contour vector points of horizontal, vertical, and diagonal segments using opencv'schain_approx_none.

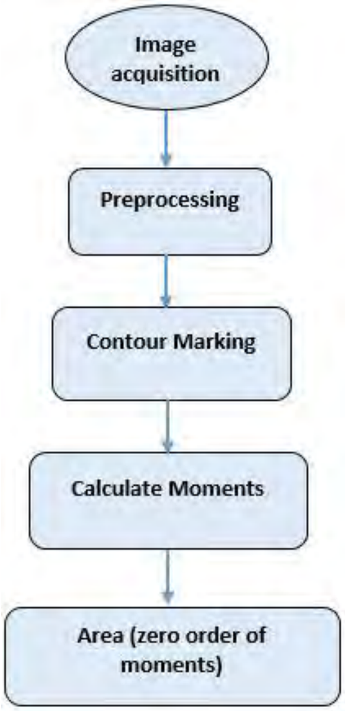


Fig. 3.3.4 Flowchart of area measurement

Now, let consider $I(x,y)$ is the image,moment of image can be calculated using following equation:

$$M_{i,j} = \sum_x \sum_y x^i y^j I(x,y) \tag{23}$$

In the cells' binary image, the zeroth moment ($M_{0,0}$) is the Area.

$A = M_{0,0}$ (16)

As the binary image's pixels are 1 and 0, if x and y are 0 that means for every white pixels, a '1' will be summed to moment to calculate the area

$$A = \sum_{x=0}^{\omega} \sum_{y=0}^{h} x^0 y^0 f(x,y) \text{(17)}$$

However, for our cells image areas that are lower than 1100 pixels are counted as normal cells and rest are counted as cancer cells.

3.4 Clump Detection Based on Cell's Center and Distance Measurement:

We proposed to mark the center of the cells and measure the positions and find out the lowest three distances from one another to reduce error, based on the values of the distances. Hence find out if a clump of cell is present. This system of center measuring is also based on the calculation of the moments of the image. Gaussian smoothing is applied just before the median filtering during pre-processing. Both Ostu algorithm and canny edge detection algorithm are applied to analysis severally. To split the touching or overlapping cells, watershed algorithm is used based ride line after getting value from distance transformation.

Then the cells are contoured. Contour contains the coordinates of cells that have the same outline intensity with fewer number of vertices. Contour approximate method is the implementation of Douglas-PeuckerAlgorithm, it stores all contour vector points of horizontal, vertical, and diagonal segments with opencv'schain_approx_none method. Moments [26] that are particular weighted average of the image pixels' intensities are calculated using Discrete Green's Theorem and the boundaries can be drawn using drawcontour function of opencv.

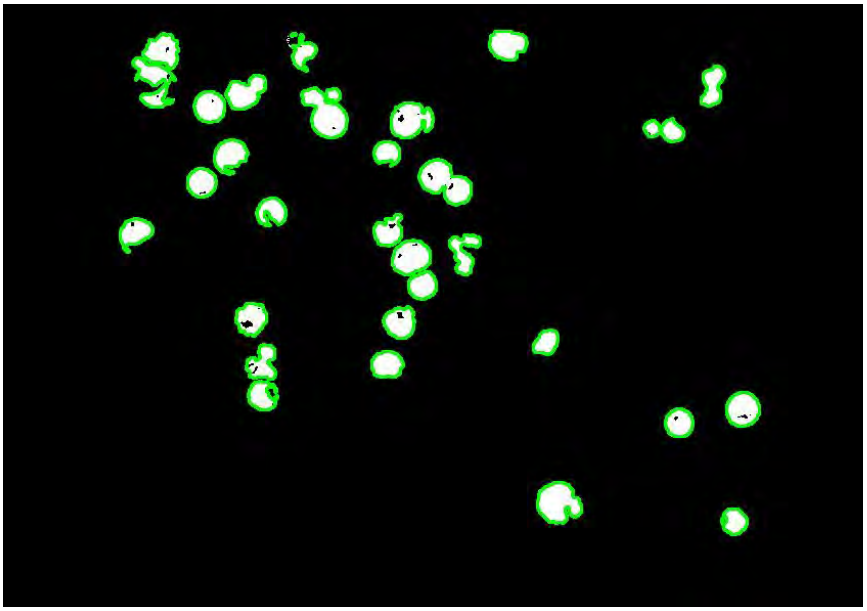


Fig. 3.4.1 Contour marked and drawn

Measuring area by moment calculation is implemented in 3.3.2,thendivide each by the number of pixels that is the zeroth moment (m00)

$$m_{10} = \frac{Sum_x}{m_{00}} \quad and \quad m_{01} = \frac{Sum_y}{m_{00}}$$

$$\text{Center} = \left[\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right]$$

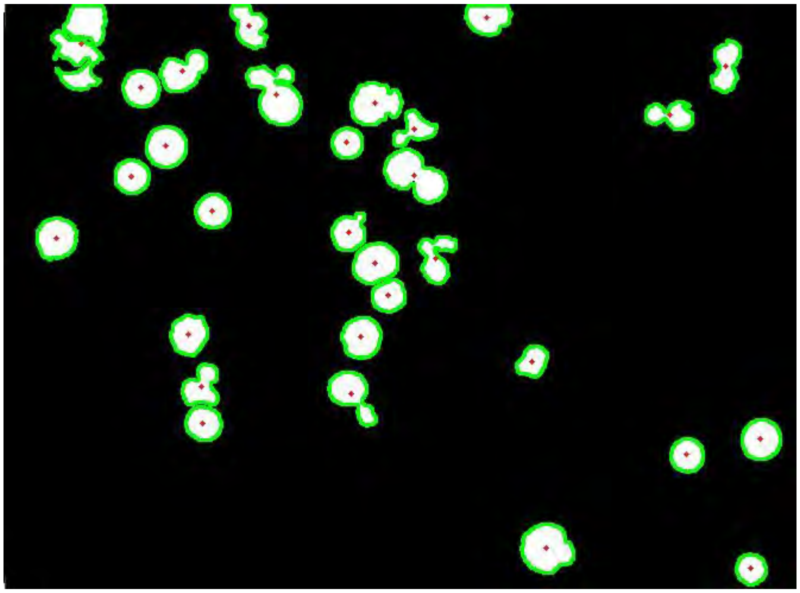


Fig. 3.4.2 Cancer Cells with marked center

After that, the distance from one center point to other is measured using distance formula.

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (24)$$

To reduce error, three of the lowest distances from each cells' center point is measured. Then if the distances are under a certain specified value, it will be counted as cancer.

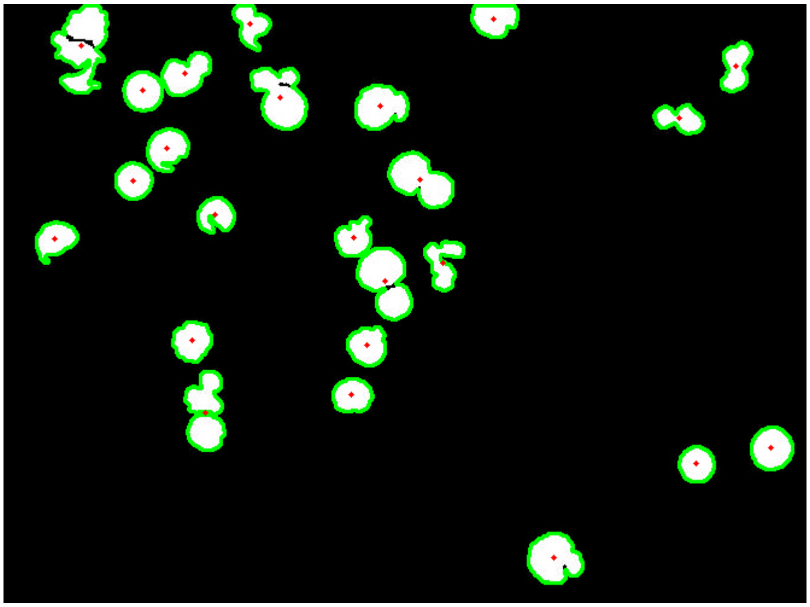


Fig. 3.4.3 Center Marked by OTSU thresh Algorithm

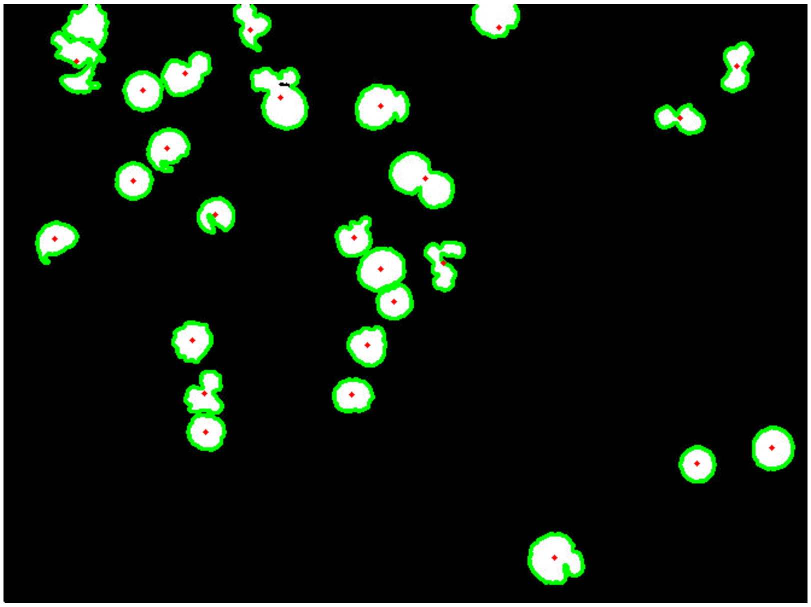


Fig.3.4.4 Center Marked for Canny Edge Detecting Algorithm

Chapter 4

Results and Limitations

4.1 Results

The proposed system is able to detect and count cells, measure areas and center marking with distance calculation . By the cell counting approach , the system found larger lymphoma cells on different area opening .

Table 1: Counted Cancer cell by trial and error method

Area (in Pixel)	Counted Cancer Cell
800	33
850	33
900	33
950	33
1000	33
1050	33
1100	31
1150	30
1200	28
1250	28
1300	28
1350	26
1400	26
1450	24
1500	23

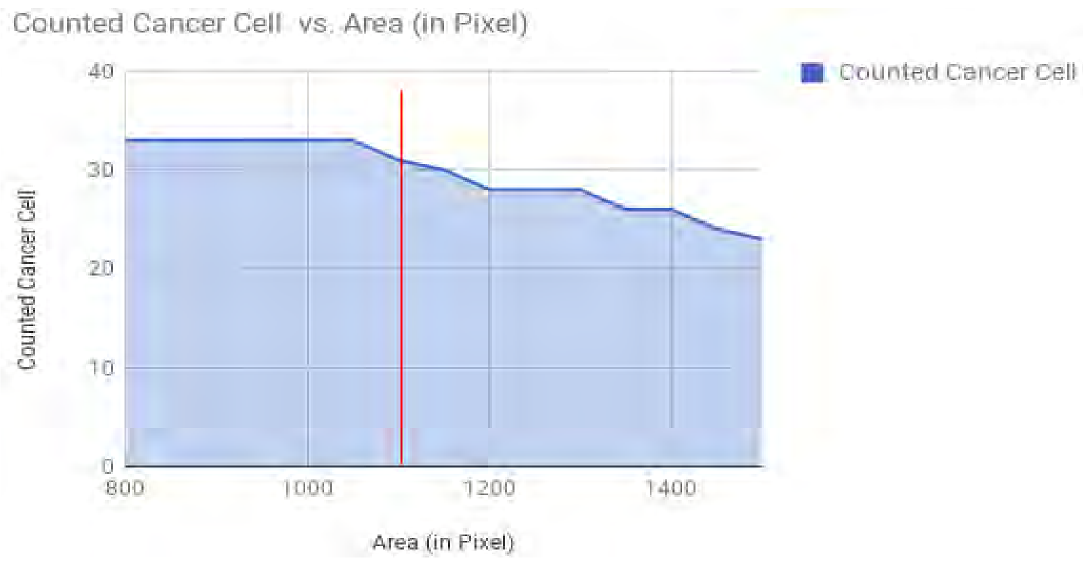


Fig. 4.1 Cancer Cell vs. Area

Here, the system gets total 96 cells and 31 cancer cell when area is more than 1100 pixel. In the Fig 4.1 , it seems 800 to 1050 pixel area there was total 33 cells and after 1100 pixels area there was 31 cell, and by trail and error method .

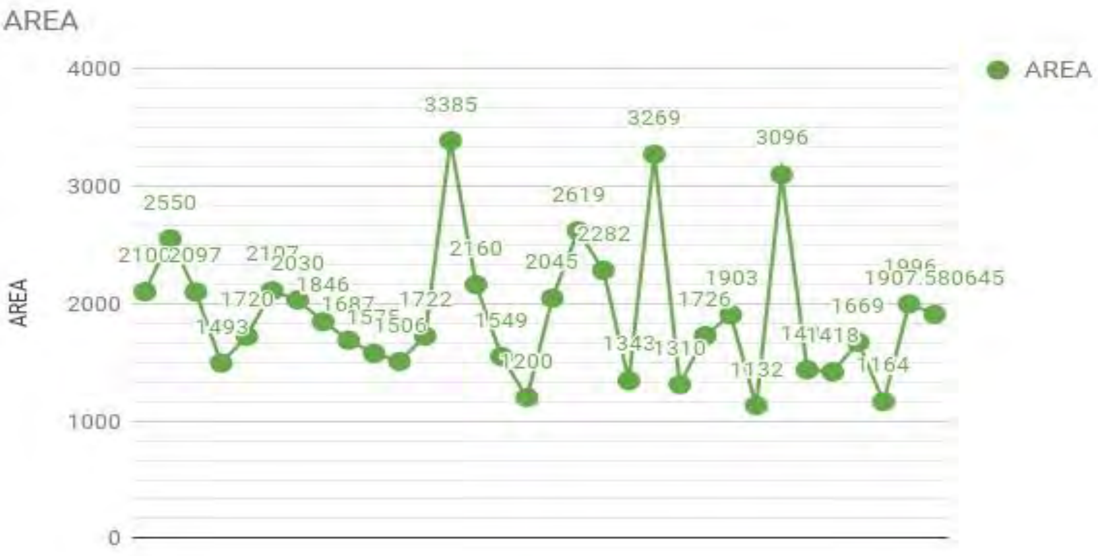


Fig. 4.2 Cancer Cell area

In the 31 cancer cells the range is from 1132 to 3385 pixels (Fig4.2) , and Fig(4.3) the systems plotted and shows that 65 cells area are under the red line, and 31 cells area above red line. The cancer cells are in quiet free from and in larger size where normal cells are much smaller in size

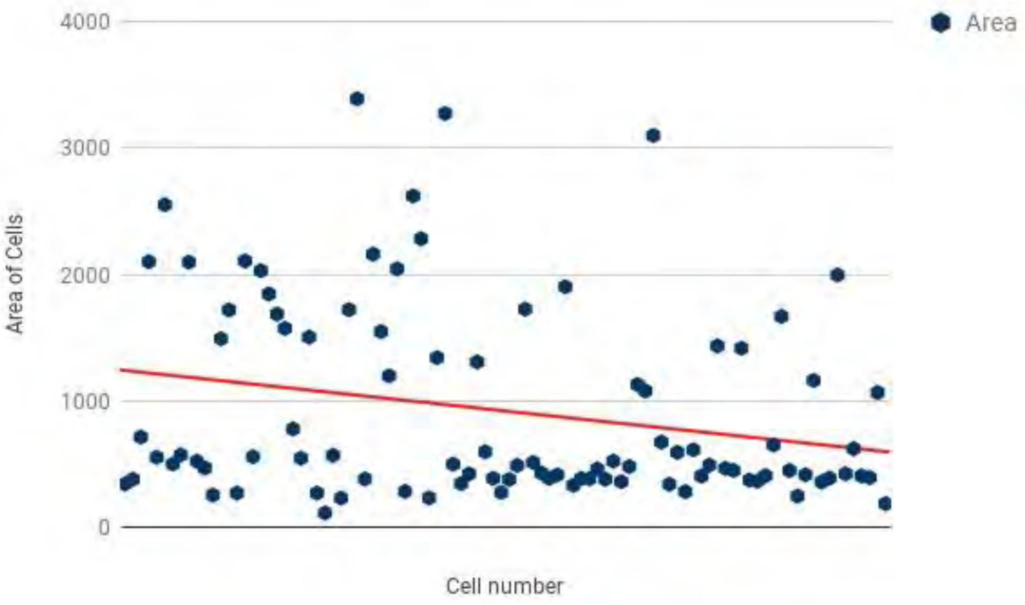


Fig. 4.3 Area of all cells in graph (points above red line are cancer cell)

By applying clump detection approach, the centers of the all cells (Table 2) and center of only cancer cells (Table 3)

Table 2 : Center of All Cells

Counted cells	Center [x,y]
1	[7.36023054755043,343.850144092219]
2	[28.2441558441558,310.961038961039]
3	[35.9958217270195,205.472144846797]
4	[65.6976190476190,298.315714285714]
5	[67.6845878136201,616.125448028674]
6	[91.8101960784314,74.3952941176471]
7	[78.1089108910891,524.702970297030]
8	[84.1421143847487,687.764298093588]
9	[104.398187887458,24.7715784453982]
10	[132.726755218216,610.153700189753]
11	[144.016877637131,188.430379746835]
12	[141.392307692308,46.2653846153846]
13	[161.191560616209,220.291359678500]

14	[172.962790697674,109.136046511628]
15	[168.514492753623,185.902173913043]
16	[204.710963455150,182.563834836260]
17	[202.955595026643,475.136767317940]
18	[225.766502463054,88.0507389162562]
19	[233.534127843987,420.101841820152]
20	[249.185536455246,486.125666864256]
21	[251.528888888889,532.521904761905]
22	[245.621483375959,355.800511508951]
23	[247.978221415608,565.898366606171]
24	[263.533200531209,264.171314741036]
25	[254.797101449275,142.391304347826]
26	[255.891666666667,446.875000000000]
27	[276.155052264808,228.426829268293]
28	[274.586497890295,484.848101265823]
29	[309.038908246225,29.3821138211382]
30	[343.451107828656,116.837518463811]
31	[417.917312661499,325.351421188630]
32	[439.091666666667,496.308796296296]
33	[435.169786959329,290.913492575855]
34	[432.999166666667,177.995000000000]
35	[450.935941320293,424.034718826406]
36	[436.916955017301,58.4152249134948]
37	[467.898816342115,128.172584956090]
38	[468.482909728309,329.747151621385]
39	[462.672268907563,23.7268907563025]
40	[485.414743112435,371.345495160089]
41	[518.569287243805,219.532884674212]
42	[492.494047619048,641.210317460318]
43	[491.971509971510,249.920227920228]
44	[498.717289719626,445.897196261682]
45	[517.683969465649,160.845038167939]

46	[512.520798668885,109.316139767055]
47	[522.826530612245,431.147959183674]
48	[527.092526690391,8.45551601423488]
49	[529.054687500000,70.6015625000000]
50	[535.093117408907,188.038461538462]
51	[545.294322132097,323.701042873696]
52	[536.897683397683,408.837837837838]
53	[549.248275862069,134.937931034483]
54	[578.068877551020,130.724489795918]
55	[581.486873508353,349.365155131265]
56	[610.276931161324,18.7241198108250]
57	[592.756676557864,319.729970326410]
58	[604.918158567775,195.179028132992]
59	[607.677835051546,356.378865979381]
60	[614.036480686695,286.982832618026]
61	[635.384415584416,689.462337662338]

62	[647.607954545455,385.282196969697]
63	[644.874316939891,620.448087431694]
64	[657.041237113402,63.4742268041237]
65	[666.686395759717,454.436395759717]
66	[669.569316081331,600.313308687616]
67	[684.578811369509,690.206072351421]
68	[672.486725663717,221.861356932153]
69	[698.791304347826,60]
70	[715.725293132328,22.2311557788945]
71	[726.351916376307,736.351916376307]
72	[728.367909238250,222.089141004862]
73	[752.968292682927,27.8536585365854]
74	[798.786290322581,58.7842741935484]
75	[841.173398328691,143.049442896936]
76	[841.719745222930,252.987261146497]
77	[841.848351648352,361.613186813187]

78	[861.739069111425,572.722849083216]
79	[857.445910290238,37.4722955145119]
80	[863.589189189189,82.5000000000000]
81	[881.518159806295,48.0145278450363]
82	[884.978625954199,249.111450381679]
83	[910.844817255842,80.9910125823847]
84	[919.006607929515,16.1541850220264]
85	[920.115079365079,543.718253968254]
86	[932.221428571429,340.545238095238]
87	[942.439862542955,716.563573883162]
88	[938.728021978022,223.670329670330]
89	[940.187817258883,44.6852791878173]
90	[955.164328657315,552.757515030060]
91	[955.670560747664,19.0257009345794]
92	[964.718849840256,250.226837060703]
93	[969.885922330097,343.483009708738]
94	[974.520100502513,111.939698492462]
95	[987.245088868101,400.219831618335]
96	[996.134715025907,239.466321243523]

Table 3 :Center of cancer cells

Counted Cancer Cell no.	Center [x,y]
1	[65.6976190476190,298.315714285714]
2	[91.8101960784314,74.3952941176471]
3	[104.398187887458,24.7715784453982]
4	[161.191560616209,220.291359678500]
5	[172.962790697674,109.136046511628]
6	[204.710963455150,182.563834836260]
7	[225.766502463054,88.0507389162562]
8	[233.534127843987,420.101841820152]
9	[249.185536455246,486.125666864256]
10	[251.528888888889,532.521904761905]
11	[263.533200531209,264.171314741036]
12	[309.038908246225,29.3821138211382]
13	[343.451107828656,116.837518463811]
14	[439.091666666667,496.308796296296]
15	[435.169786959329,290.913492575855]
16	[432.999166666667,177.995000000000]
17	[450.935941320293,424.034718826406]
18	[467.898816342115,128.172584956090]
19	[468.482909728309,329.747151621385]
20	[485.414743112435,371.345495160089]
21	[518.569287243805,219.532884674212]
22	[517.683969465649,160.845038167939]
23	[545.294322132097,323.701042873696]
24	[610.276931161324,18.7241198108250]
25	[666.686395759717,454.436395759717]
26	[684.578811369509,690.206072351421]
27	[841.173398328691,143.049442896936]
28	[861.739069111425,572.722849083216]
29	[910.844817255842,80.9910125823847]
30	[942.439862542955,716.563573883162]
31	[955.164328657315,552.757515030060]

The system is run using Matlab and OpenCV with Python . By running 3time per approach , the system gets the time complexity (Fig4.5)

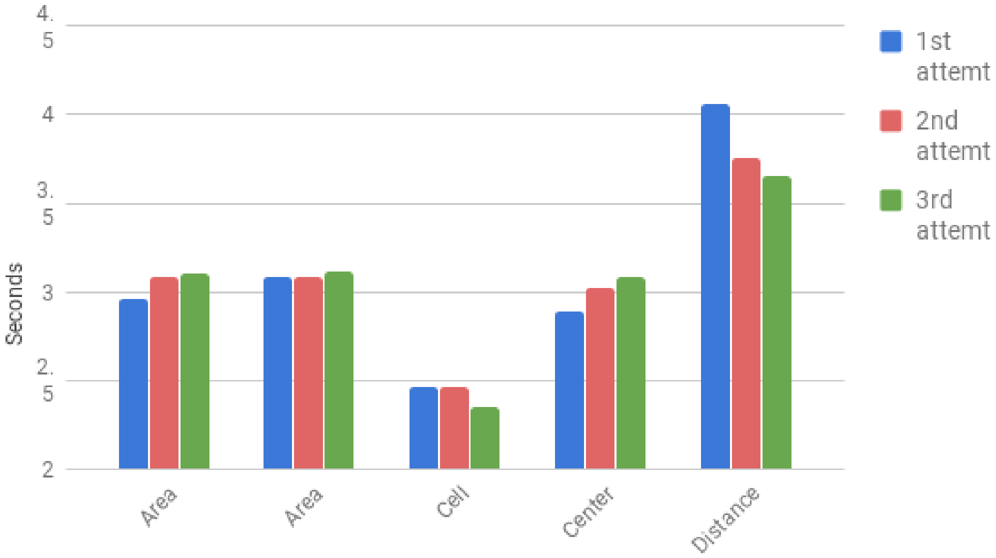


Fig 4.5 : Time complexity of the system

Chapter 5
Conclusion

5.1 Concluding Remarks

Every year millions of people are affected by cancer all over the world. And a significant percentage of these people die because there is no solid cure to the type of cancer that affected them. Numerous scientific communities are constantly researching on different grounds of cancer to figure out possible defined cures. In this research, the proposed model effectively detects cancer cells autonomously, by means of *cell counting, area measurement of cells, and clump detection.* The system is flexible enough to detect various types of cancer based on several manual inputs set by the user. As a result a wide variety of cancer can be detected respectively.

5.2 Future Works

This article has presented a novel approach to detect cancer cells from microscope images of tissue sample slides. The system is already due to be tested with a large number of real data set collected from various cancer research centers.

A user friendly GUI will be implemented soon after trial and error process with the large set of real data.

References

[1]http://lymphomapictures.org/p/37/non-hodgkin-lymphoma/picture-37

[2] Kumar R., Srivastava R., Srivastava S. “Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features” Proc of Journal of Medical Engineering, Volume 2015 (2015), Article ID 457906, 14 pages.

[3] Patil,G.Bhagyashri. , Jain, Sanjeev.N. (2014), “Cancer Cells Detection Using Digital Processing Methods” International Journal of Latest Trends in Engineering and Technology.

[4] Ramin M., Ahmadvand, P., Sepas-Moghaddam, A.Dehshibi, M.M., “Counting Number of Cells in Immunocytochemical Images using Genetic Algorithm,” 12th International Conference on Hybrid Intelligent Systems, Dec. 2012. pp. 185-190.

[5] Dahle J. , Kakar M. , Steen H.B. , Kaalhus O., “Automated counting of mammalian cell colonies by means of a flatbed scanner and image processing”, 29 June 2004,ISAC,10.1002/cyto.a.20038.

[6] Venkatalaksmi B. and Thilagavathi K., “Automatic Red Blood Cell Counting Using Hough Transform,” Proc. of 2013 IEEE Conference on Information and Communication Technology, Apr. 2013, pp. 267-271.

[7] Bergmeir C. , Silvente M. G., and Benítez J. M., “Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework,” Computer Methods and Programs in Biomedicine, vol. 107, no. 3, pp. 497–512, 2012.

[8] Carlos A. B. Mello, Wellington P. dos Santos, Marco A. B. Rodrigues, Ana LciaB.Candeias, Cristine M. G. Gusmao ,“Image Segmentation of Ovitraps for Automatic Counting of AedesAegypti Eggs,” 30th Annual International IEEE EMBS Conf. Vancouver, British Columbia, Canada, Aug. 2008. pp. 3103-3106.

[9]Gonzalez, R.C. & Woods, R.E. (2006). Digital Image Processing. NJ,USA:Prentice-Hall,Inc.Upper Saddle River.

[10]Introduction to image processing. Retrieved from https://sisu.ut.ee/imageprocessing/book/1

[11]Efford, N. (2000). *Digital Image Processing: A Practical Introduction Using Java*. Pearson Education.

[12]Williams, J. (2013). *Morphological Image Processing*. Retrieved from: https://www.slideshare.net/Johnrebel999/morphological-image-processing-22899372

[13] Amon, G.(2012, April 9). Image Segmentation Digital Signal Processing. Retrieved from: https://www.slideshare.net/gichelleamon/image-segmentation-ppt

[14] A history of medical imaging (2017). Retrieved from: http://www.infinityugent.be/research-development/a-history-of-medical-imaging

[15] R. Boyle and R. Thomas *Computer Vision: A First Course*, Blackwell Scientific Publications, 1988, page. 32 - 34.

[16]E. Davies *Machine Vision: Theory, Algorithms and Practicalities*, Academic Press, 1990, Chap. 3.

[17] Gonzalez,R.C and Woods,R *Digital Image Processing*, Addison Wesley, 1992, pp 414 – 428

[18] B. Jähne, H. Scharr, and S. Körkel. Principles of filter design. In Handbook of Computer Vision and Applications. Academic Press, 1999.

[19]A. EL Allaoui, M. Merzougui, M. Nasri, M. EL Hitmy and H. Ouariachi. Evolutionary Image Segmentation by Pixel Classification Application to Medical Images. IJCIIS International Journal of Computational Intelligence and Information Security*, ISSN: 1837-7823*, Vol. 2, No. 3 pp. 12-24. March 2011.

[20] A. Preetkaur, A. Verma, "The Marker-Based Watershed Segmentation –A Review" International Journal of Engineering and Innovative Technology (IJEIT), Volume 3, Issue 3, September 2013.

[21] Ravi S and A M Khan, "Morphological Operations for Image Processing: Understanding and its Applications", Proc. 2nd National Conference on VLSI, Signal processing & Communications NCVSComs-2013, 11 -12th Dec. 2013.

[22]Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Systems, Man and Cybernetics, 9(1):62-66, 1979.

[23]Prasad, Dilip K.; Quek, Chai; Leung, Maylor K.H.; Cho, Siu-Yeung (2011). *A parameter independent line fitting method*. 1st IAPR Asian Conference on Pattern Recognition (ACPR 2011), Beijing, China, 28-30 Nov.

[24]T.Avery and G. Berlin *Fundamentals of Remote Sensing and Airphoto Interpretation*, Maxwell Macmillan International, 1985, Chap. 15.

[25] R.Gonzalez and R. Woods *Digital Image Processing*, Addison-Wesley Publishing Company, 1992, Chap. 2.

[26]Rosebrock, A. (2016, February 1). OpenCV center of contour. Retrieved from http://www.pyimagesearch.com/2016/02/01/opencv-center-of-contour/