



# Clustering Deviations in Medical Images using Hierarchical Clustering and Shortest Path Algorithms

M. Vaishnavi<sup>1</sup>, P. Akhila<sup>2</sup>, N. Jyothsna<sup>3</sup>, P. Alekhya<sup>4</sup>, T. Sudhir<sup>5</sup>  
B.Tech Student<sup>1, 2, 3, 4</sup>, Professor<sup>5</sup>

Department of Computer Science and Engineering  
Vasireddy Venkatadri Institute of Technology, AP, India

## Abstract:

Most of the patient diagnosis revolves around in identifying abnormalities in their respective medical images. These images are of various types, likely Ultrasound, CT Scan, MRI and microscopic images like bio-chemical slides, micro-biological slides & pathological slides. Few abnormalities are fractures, bad cells in blood, tumors, fungal identification etc. Finding the abnormal portions in these images needs expertise by the physician; this apt identification promotes and guarantees healthy medication by the physician or surgeon to patient. In medical microscopic images normal portions and abnormal portions are mixed together. None of the abnormal portions are related to abnormal and normal portions of image i.e. deviations are scattered among normal portions of image. These deviations are not present in some portions for specific area in the images. None of these deviations are overlapped nor can be grouped together into a single portion physically in the image. Deviations can be isolated along with normal portions of images. Identifying such deviations partially comes under clustering. This project identifies deviations in Medical Microscopic images. These deviations can be identified visually which reveals about the presence of deviation but to know the percentage of deviation in a sample image is imperative. In-order to achieve this all deviations must be connected. This project connects all deviations using Shortest Path algorithm and clusters using Hierarchical Clustering algorithms.

**Keywords:** Cluster, Deviations, Hierarchical Clustering, Shortest Path

## I. INTRODUCTION

Medical Imaging is the technique and process of creating virtual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues. Different types of medical images are Scanned Images and Microscopic Images. Scanned images include MRI scan, CT Scan, PET, X-Ray, Ultrasound.

### Types of Medical images

#### Magnetic Resonance Imaging (MRI)

MRI is a medical imaging Technology that uses radio waves and a magnetic field to create detailed images of organs and tissues. MRI is used to evaluate blood vessels, Abnormal tissues, Bones and Joints, Spinal injuries etc.

#### Computed Tomography (CT)

Computed Tomography (CT), is a medical imaging method that combines multiple X-ray projections taken from different angles to produce a detailed cross-sectional images of areas inside the body. CT images allow doctors to get very precise, 3D views of certain parts of the body. CT is used to evaluate presence, size, location of tumors, Bone injuries, Organs in chest, abdomen etc.

#### Positron Emission Tomography (PET)

PET is a nuclear imaging technique that provides physicians with information about how tissues and organs are functioning. PET is used to evaluate Neurological diseases such as Alzheimer's and Multiple Sclerosis, Cancer etc.

#### Ultrasound

Diagnostic ultrasound, also known as Medical Sonography, uses high frequency sound waves to create images of parts inside the

body. It is used to evaluate Pregnancy, Abnormalities of heart and blood vessels etc.

#### X-Ray

X-rays use ionizing radiation to produce images of a person's internal structure by sending X-ray beams through the body. X-ray images are used to evaluate Broken bones, Cavities, Lungs etc.

#### Microscopic Images

Microscopic images are used in medical diagnosis. They were made with a very high resolution of 4096 x 2160 pixels(4K), which is four times as many pixels as High-Definition image(1920 x 1080 pixels).

#### The various types of microscopic images are:

- Biochemical sides
- Pathological slides
- Micro-biological slides

Finding the abnormal portions in these images needs expertise by the physician; This apt identification promotes and ensures healthy medication by the physician to patient.

## II. RESEARCH ELABORATIONS

### Anaemia

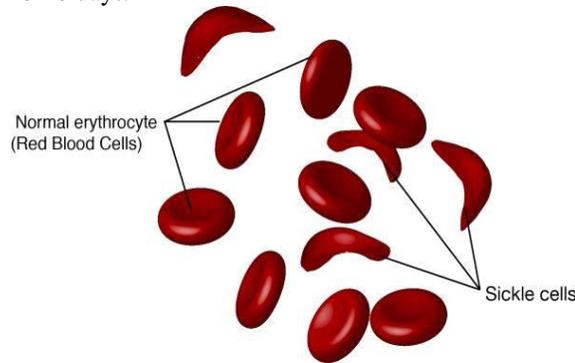
The cellular part of blood molecule contains several different cell types like red blood cells, white blood cells and platelets. Anemia is the most common disorder of the blood.

“Anemia”, the name is derivative from the ancient Greek word anemia, which means “Lack of Blood”. It is possible because of reduction in Red Blood Cells (RBCs) or resulting in lesser than normal quantity of haemoglobin in the blood. However, it can also include decreased oxygen-binding ability of each haemoglobin molecule due to deformity or lack in numerical development [2].

Anemia is actually a sign of a disease process rather than being a disease itself. It can be either classified as acute or chronic. In chronic anemia, symptoms typically begin slowly and progress gradually; whereas in acute anemia, symptoms can be abrupt and more distressing. Among many factors, both nutritional (like vitamins and mineral deficiencies) and non-nutritional (like infection and haemoglobinopathies), that contribute to the onset of anemia; Iron Insufficiency and malaria plays a significant role. For men, anemia is typically defined as hemoglobin level of less than 13.5 g/dl and in women as hemoglobin of less than 12.0 g/dl.

**Sickle Cell Anemia**

Sickle cell disease is an inherited blood disorder. Red blood cells contain hemoglobin which carries the oxygen in the blood. They are round and flexible which enables them to travel through small blood vessels to deliver oxygen to all parts of the body. These cells normally live about 120 days. Sickle cell disease causes red blood cells to become crescent shaped, which causes them to break apart easily. This ultimately causes anemia because the cells only live 10-20 days.

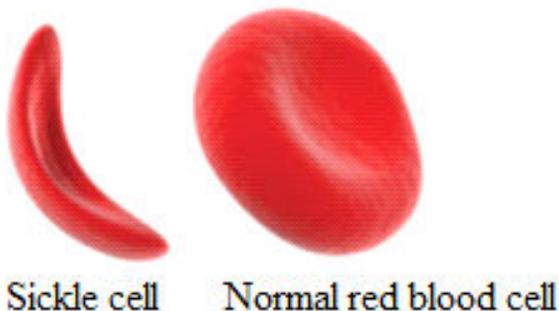


**Figure.1. Sickle cells and Normal Blood Cells**

**Characterization of Sickle-Cell Anemia:**

When the abnormal sickle-shaped cells in the blood are identified, a diagnosis is made.

- Normal Cell: Normally blood cells are round and flexible and flow easily through blood vessels.
- Sickle Cell: In sickle cell disease, certain red blood cells become crescent-shaped. These abnormal red blood cells, carrying abnormal hemoglobin known as hemoglobin S, are fragile.

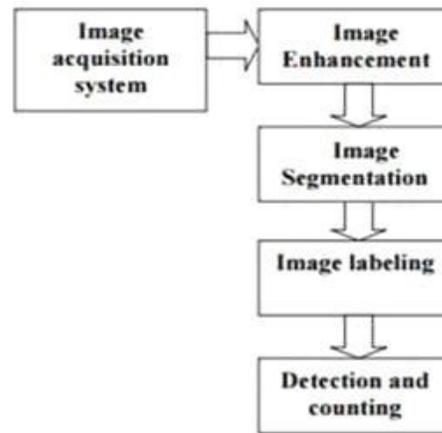


**Figure.2. Normal RBC and Sickle Shaped RBC**

In human health, blood health is very important, and the early diagnosis of diseases is necessary to prevent death and the worsening of disease. With the development of information technology, image-processing technology is becoming an essential and effective tool in scientific research. It is especially widely used and effective in the field of biomedical engineering. Finding a faster, more efficient methods to diagnose this kind of disease by using image segmentation so that sickle cells can be detected in blood samples is an essential part of detecting this disease. The goal of this thesis is to demonstrate how image processing technology can be useful in the detection of sickle cell disease.

**III. EXISTING SYSTEM**

There have been many studies about sickle blood cell detection. One of them is the “Detection of Abnormal Blood Cells Using Image Processing Technique” by Malher Bhatt and Shashi Probha. In this study, the authors try to detect abnormal blood cells based on form factor and using image segmentation. The form factor they used calculates the circularity shape of the cell, and if the cell not circular, it is counted as an abnormal cell.



**Figure.3. Block diagram for detecting abnormal cells**

**A. Image Acquisition**

The digital microscope is interfaced to a computer. The microscopic images are obtained as digital images. The resolution of the digital image depends on the type of digital microscope used [6].

**B. Image Enhancement**

For better segmentation of the blood cells, the imported image has to be enhanced. This improves the quality of the image by enhancing the details [6].

**C. Image Segmentation**

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. [7].

**D. Abnormal Blood Cell Detection**

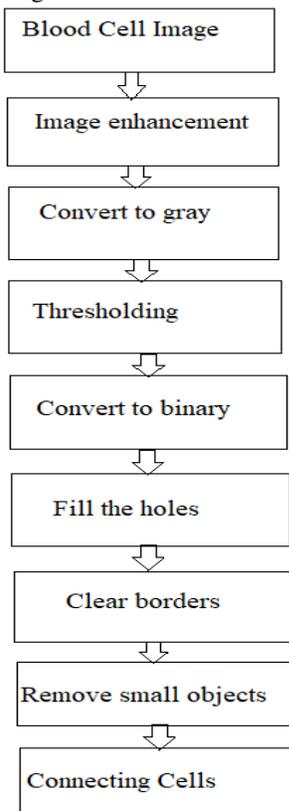
Form factor threshold is fixed for different abnormal cells. Based on the form factor, the abnormal cells are detected. [6]

Malher and Probha started with reading the image and then converted it to a grayscale image. They then used edge detection to detect the abnormal cells. Edge detection is an image segmentation technique that determines the presence of an edge or line in an image and outlines the edge in an appropriate way. The main purpose of edge detection is to simplify the image data in order to minimize the amount of data to be processed [9]. Edge detection is followed by a labeling stage and then form factor calculation [6]. Form Factor =  $4 * \text{Pi} * \text{area} / (\text{perimeter} * \text{perimeter})$  [8][9]. Here form factor is calculated and it is used to determine whether the cell is normal or sickle cell. In this the sickle cells can only be counted but cannot be connected.

## V. PROPOSED ANALYSIS

### A. Flow of Proposed Solution

The blood cell image (microscopic) is given as input and later the blood cell image is enhanced to get in the desired format. After being enhanced, these are converted to gray images using built-in functions in Matlab. After thresholding, the gray scale image is converted to binary image using the function `im2bw()`. The binary image thus obtained may contain some holes. In order to get the accurate output, we need to fill these holes. This can be done by using the Matlab function `imfill()`. The objects that are present at the borders are removed using the function `imclearborder()`. After clearing borders, small objects are removed. This will be done by detecting the objects with area that are less than a specific value. And this can be done using the function `Bwareaopen()`. After deleting the small objects, cells will be connected. This requires the detection of every unit of the cell and a function `pic2points` is defined which converts the image coordinates into graph coordinates and further operations like connecting is done here. `knnsearch(X,Y,Name,Value)` is used to accomplish this task. At the end, we get cells that are connected as output.



## VI. IMPLEMENTATION

### MATLAB environment

To implement the new algorithm, the "MATLAB" tool is used. MATLAB has an image processing toolbox, which contains all functions that are used to analyze the image such as reading, enhancement, converting from one image type to another, segmentation, labeling and more. The functions that used to run the new algorithm are described below [4].

#### A. Reading Image (Imread)

MATLAB deals with the image as a two dimensional matrix. `Imread` reads the image into MATLAB's environment as a matrix[4].

Basic Syntax

`Image=imread('file-name');`

MATLAB reads the image file "file-name" into an array image. `Imread` can read many different formats such as JPG, PNG, GIF and TIF[4].

#### B. Showing Image (Imshow)

`Imshow` displays images on the MATLAB desktop.

Basic Syntax

`Imshow(I)`

MATLAB displays the image `I` in a graphics figure, where `I` is a grayscale, RGB (truecolor), or binary image. For binary images, `imshow` displays pixels with the value 0(zero) as black and 1 as white[4].

#### C. Converting Colored Image to Gray (rgb2gray)

There are four types of image:

- 1- RGB image
- 2- Gray – Scale images
- 3- Binary image "Black and white"
- 4- Indexed images

**Rgb2gray converts the image from RGB into a gray-scale image.**

Basic Syntax

`I=rgb2gray(RGB);`

MATLAB converts the true color image RGB to the grayscale intensity image `I`. The `rgb2gray` function converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance. If you have Parallel Computing Toolbox installed, `rgb2gray` can perform this conversion on a GPU[4].

#### D. Calculating Threshold and Converting Gray Image to Binary (im2bw)

Converting the image from RGB into a binary image first requires computing the threshold of the gray image using the "graythresh" function.

Basic Syntax

`Level = graythresh(RGB);`

MATLAB computes a global threshold (Level) that can be used to convert an intensity image to a binary image with `im2bw`.

Level is a normalized intensity value that lies in the range [0, 1]. The graythresh function uses Otsu's method, which chooses the threshold to minimize the intraclass variance of the black and white pixels. Then this threshold converts the image to black and white using "im2bw".

Basic Syntax

`BW = im2bw (Grayimage, Level);`

MATLAB converts the grayscale image Grayimage to a binary image. The output image BW replaces all pixels in the input image with luminance greater than Level with the value 1 (white) and replaces all other pixels with the value 0 (black). It specifies a Level in the range [0,1]. This range is relative to the signal levels possible for the image's class.

Therefore, a Level value of 0.5 is midway between black and white, regardless of class. If you do not specify Level, im2bw uses the value 0.5[4].

### E. Filling Images Holes (Imfill)

Imfill fills image regions and holes.

Basic Syntax

`Filled = imfill (BW, 'holes');`

MATLAB fills holes in the input binary image BW. In this syntax, a hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image [4].

### F. Clearing the border (Imclearborder)

This function clears any objects that touch the image borders.

Basic Syntax

`Cleared = imclearborder (Filled);`

MATLAB suppresses structures that are lighter than their surroundings and that are connected to the image border. Use this function to clear the image border. Filled can be a grayscale or binary image. For grayscale images, imclearborder tends to reduce the overall intensity level in addition to suppressing border structures. The output image, Cleared, is grayscale or binary, depending on the input. The default connectivity is 8 for two dimensions, 26 for three dimensions, and `conndef (ndims (Filled), 'maximal')` for higher dimensions[4].

### G. Deleting Small Objects (Bwareaopen)

This function deletes any object that's area is less than a specific value.

Basic Syntax

`Deleted = bwareaopen (Cleared, Value);`

MATLAB removes all connected components (objects) that have fewer Value pixels from the binary image Cleared, producing another binary image, Deleted. The default connectivity is 8 for two dimensions, 26 for three dimensions, and `conndef (ndims(Cleared), 'maximal')` for higher dimensions. This operation is known as an area opening[4].

### H. Connecting Cells

Connecting of all the cells requires the detection of every unit of the cell. To achieve this, we have defined a function pic2points which converts the image coordinates into graph coordinates on which we can perform the further operations like connecting. To accomplish the task, we make use of a predefined function named "knnsearch". `knnsearch(X,Y,Name,Value)` returns Idx with additional options specified using one or more name-value pair arguments using which we can specify the number of nearest neighbors to search for and the distance metric used in the search.

## VII. ANALYSIS AND RESULTS

### Sample1

#### Input Image



Figure.4. Results after reading input sample1.jpg file



Figure.5. Results after converting to gray scale

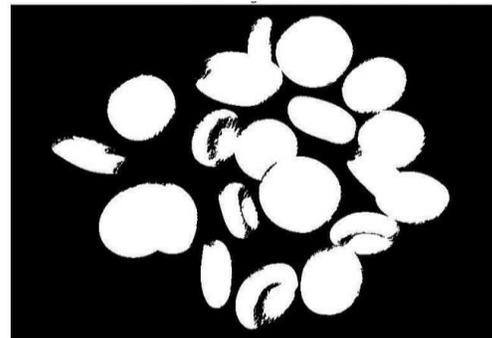


Figure.6. Results after converting to binary image

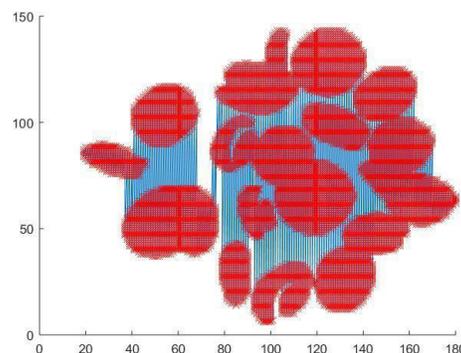
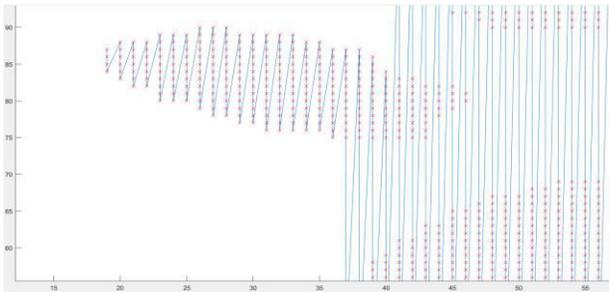


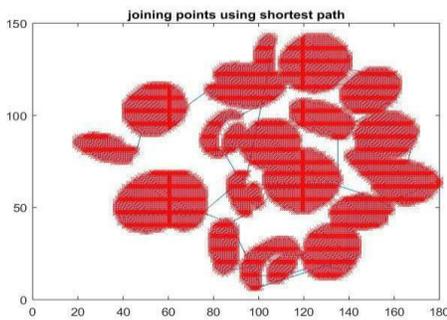
Figure.7. Results after connecting every possible unit

Thresholding is done and image is converted to binary image. The holes present are also removed using `imfill()` function which results in the above image. All the units of the cells are connected using the function `line(x,y)` which gives the above output.

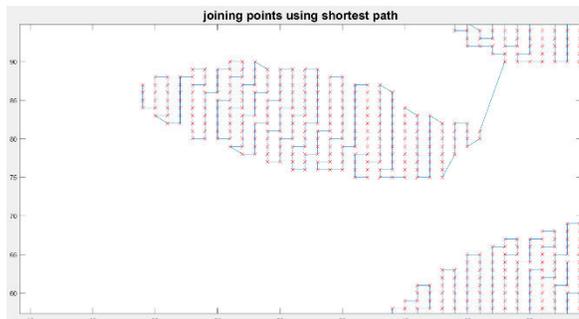


**Figure.8. Closer view of units of matrix getting connected**

The above figure shows how the points are connected using `line(x,y)` function without considering shortest paths which results in large path lengths.



**Figure.9. Results after connecting the cells using knnsearch**



**Figure.10. Closer view of results after connecting the cells using knnsearch**

Using `knnsearch` the points are connected in such a way that every point is connected to another point that is possibly nearest to it. This allows in connecting the points in a path that is shortest.

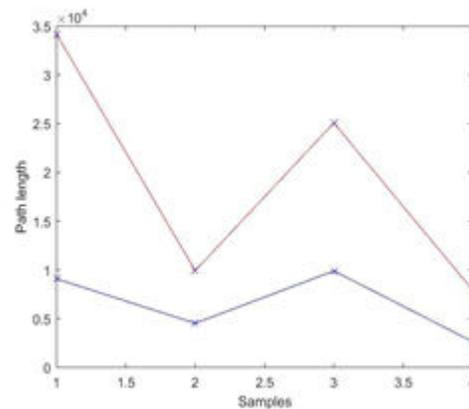
**Table.1. Comparison of complexities of different Shortest Path Algorithms**

ALGORITHMS	Time Complexity	Space Complexity
Dijkstra	$O( E + V \log V )$	$O(v^2)$
Bellman-Ford	$O(VE)$	$O(v^2)$
Floyd Warshall	$O(n^3)$	$O(n^3)$
Johnson	$O(V^2\log V + VE)$	

The algorithm that is used in this solution has a Time complexity of  $O(n^2)$  Four sample images are taken and the cells in them are joined using `line(x,y)` function and using our algorithm. The following results are obtained:

**Table.2. shows the lengths of paths for connecting the cells using `line()` function and our algorithm.**

Sample	Length when joined using <code>line()</code>	Length when joined using our algorithm
1	2.5069e+04	9.8883e+03
2	9.9243e+03	4.5201e+03
3	3.4179e+04	9.0850e+03
4	709291e+03	2.6402e+03



**Figure.11. Graph showing path lengths for different sample**  
The graph in red line indicates the length of path when joined using `line()` function and the graph in blue line indicates the length of path when joined using `knnsearch` for different samples. This clearly shows that our algorithm connects the cells in a shortest path.

### VIII. CONCLUSION

The proposed algorithm connects the cells in the medical images using `knnsearch` with more accuracy by considering each and every unit present. The points are connected in a way that the path formed by connecting them is shortest. Compared to the prior proposed solutions, this solution connects all the possible units without ignoring any unit and in a shortest path.

### IX. REFERENCES

[1]. Sickle Cell Disease Symptoms, Causes,Treatments-Web MD.(n.d.).from <http://www.webmd.com/pain-management/pain-management-sickle-cell-disease?page=2>.

[2]. Siddharth Barpanda, (May-2013), "Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear"from <http://ethesis.nitrkl.ac.in/5022/1/109EE0255.pdf>

- [3]. "Types of Medical Images" <http://www.medicalimaging.org/about-mita/medical-imaging-primer/>
- [4]. "Image Processing Functions in Matlab" <http://in.mathworks.com/help/images/functionlist.html>
- [5]. "Medical Imaging", [https://en.wikipedia.org/wiki/Medical\\_imaging](https://en.wikipedia.org/wiki/Medical_imaging)
- [6]. M. Bhatt1 & S. Prabha, (Jan- June 2015), "Detection of Abnormal Blood Cells Using Image Processing Techniques", IJEEE, Vol 07, Issue 01, pp.89-93.
- [7]. Manjula S1, Rashmi M.J2 & Varsha.D, (April 2015), "Sickle Cell Detection Using Marker controlled Watershed Segmentation", International Journal of Advance Research In Science And Engineering (IJARSE), Vol. No.4, Special Issue (01), ISSN-2319-8354(E).
- [8]. Deepika N. Patil & Uday P. Khot, (September, 2015), "Image Processing Based Abnormal Blood Cells Detection," [www.ijtra.com](http://www.ijtra.com) Special Issue 31, PP. 37-43.
- [9]. Rakshit, P., & Bhowmik, K. (2013). Detection of Abnormal Findings in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing. *Procedia Technology*, 10, 28-36. doi:10.1016/j.protcy.2013.12.333.
- [10]. Kholoud Alotaibi, " Sickle Blood Cell Detection Based on Image Segmentation",Theses and Dissertations, South Dakota State University, 2016