ISSN No:-2456 –2165

Detection of Sickle Cell Anemia and Thalassemia using Image Processing Techniques

Lavanya T H
Assistant professor, Dept. of IS&E
Channabasaveshwara Institute of Technology,
Gubbi, Tumkur

Sushritha S
Assistant Professor, Dept. of CS&E
Channabasaveshwara Institute of Technology,
Gubbi, Tumkur

Abstract— Sickle cell anemia and Thalassemia are lifelong genetic disorders characterized by Red Blood Cells that assume an abnormal sickle shape, abnormal hemoglobin structure. Sickling and structural change of hemoglobin decreases the flexibility of the cell and results in risk of various complications. This alters the cells morphology and this result can be used for classification. The proposed system takes a magnified image of the blood smear and applies various image processing techniques to detect the different disorders of red blood cell. White blood cells, platelets will automatically be removed from the image; this reduces the region of interest.

Keywords—Image Processing, Edge detection, Shape Detection, Sickle Cell, Thalassemia, K-means, Ellipse Detector.

I. INTRODUCTION

Red blood cells (RBCs), also called erythrocytes, are the most common type of blood cell. The main function of RBC is delivering oxygen (O2) to the body tissues via the blood flow through the system. They take up oxygen in the lungs or gills and release it into tissues while squeezing through the body's capillaries. The cytoplasm of erythrocytes is rich in hemoglobin, an iron-containing biomolecule that can bind oxygen and is responsible for the red color of the cells. The cell membrane is composed of proteins and lipids, and this structure provides properties essential for physiological cell function such as deformability and stability while traversing the circulatory system and specifically the capillary network.

Sickle cell anemia is a serious inherited blood disorder where the red blood cells, which carry oxygen around the body, develop abnormally. Normal red blood cells are flexible and disc shaped, but in sickle cell anemia they can become rigid and shaped like a crescent (or sickle). The sickle shaped cells contain defective haemoglobin, the iron rich protein that enables red blood cells to carry oxygen from your lungs to the rest of the body. The abnormal cells are also unable to move around as easily as normal shaped cells and can block blood vessels, resulting in tissue and organ damage and episodes of severe pain. Such episodes are known as a sickle cell. The abnormal blood cells also have a shorter lifespan and aren't replaced as quickly as normal blood cells. This leads to a shortage of red blood cells, known as anemia.

Thalassemia is a blood disorder passed down through families (inherited) in which the body makes an abnormal form of hemoglobin. Hemoglobin is the protein in red blood cells that carries oxygen. The disorder results in large numbers of red blood cells being destroyed, which leads to anemia.
There are two main types of thalassemia:

- Alpha thalassemia occurs when a gene or genes related to the alpha globin protein are missing or changed (mutated).
- Beta thalassemia occurs when similar gene defects affect production of the beta globin protein.

II. LITERATURE SURVEY

A. Edge Detection of Sickle Cell in Red Blood Cell

Authors: Aruna N.S and Hariharan S
By detecting lowest, highest and mean radius of each cell and comparing it with standard cell size we can classify the sickle and normal cell. For this process in this paper they are used edge detection techniques to mark sickle cells with red circles. In clinical method they find normal, sickle and other abnormal cells by using microscope but it is difficult to find the overlapped cells in manual method. In this automated method overlapped and incomplete blood cells are detected first and then they are classified into normal or abnormal cells based on the shape of each cell extracted from the microscopic image using different edge finding algorithms.[3]
Disadvantage:

- Requires more computation to find overlapped cells

B. Red Blood Cell Estimation Using Hough Transformation Technique

Authors: Nasrul Humaimi Mahmood and Muhammad Asraf Mansor
By calculating form factor based on morphological shape it classifies the RBC cells. It uses the circular hough transformation technique to find circular shape. It compares the circle with standard form factor range from 0.5 to 1; if the

cell is within the standard range then it is classified as normal or else classified as abnormal.[4]

Disadvantage:

- This method of abnormal cell detection did not address half-cell counting.

C. Automated Diagnosis of Thalassemia Based on Data Mining Classifiers

Authors: Eyad H. Elshami and Alaa M. Alhalees

Thalassemia is a genetic disease that is commonly found in many parts of the world. It leads to death in most of its major cases so we must control it by determining the persons who trait the thalassemia genes. Complete Blood Control (CBC) is the first and the simplest test which can narrow to the existence of thalassemia. This paper presents an investigation for thalassemia existence by using data mining classifiers depending on CBC. Three data mining classifiers were used in this investigation. Each of the classifiers used to differentiate between thalassemia traits patients- with its different levels-: iron deficiency patients, normal persons, and the patient who suffer from other blood diseases. [5]

D. Thalassemia Classification by Neural Networks and Genetic Programming

Authors: Wongseree W. Chaiyaratana, N. Vichittumaros, K. Winichagoon

Wongseree et.al [6] investigated thalassemia classification by using a neural network and a decision tree, which is evolved by genetic programming, in thalassemia classification. The aim is to differentiate between thalassemic patients, persons with thalassemia trait and normal subjects by inspecting characteristics of red blood cells, reticulocytes and platelets. But they need in the proposed model more blood testing like Platelet and Reticulocyte.[7]

III. PROPOSED SYSTEM

1. Sickle Cell Anemia Detection

A. Ellipse Detection Using Deferential Evalution

a). Data Preprocessing. In order to detect ellipse shapes, candidate images must be preprocessed first by an edge detection algorithm which yields an edge map image. Then, the (x_i, y_i) coordinates for each edge pixel $p\ i$ are stored inside the edge vector $P = \{p_1, p_2, \dots p_{N_p}\}$ with Np being the total number of edge pixels.

b). Individual Representation. Just as a line requires two points to completely define its characteristics, an ellipse is defined by five points. Therefore, each candidate solution E (ellipse candidate) considers five edge points to represent an individual. Under such representation, edge points are selected following a random positional index within the edge array P. This procedure will encode a candidate solution as the ellipse that passes through five points p1,p2,p3,p4 and p5 Thus, by substituting the coordinates of each point of E into (5), we gather a set of five simultaneous equations which are linear in the five unknown parameters a, b, f, g, and h:

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + 1 = 0.$$

Considering the configuration of the edge points, the ellipse center (x_0, y_0) the radius maximum ($r$max), the radius minimum ($r$min), and the ellipse orientation (θ) can be calculated as follows:

$$x_0 = \frac{hf - bg}{C},$$

$$y_0 = \frac{gh - af}{C},$$

$$r_{\max} = \sqrt{\frac{-2\Delta}{C\,(a + b - R)}},$$

$$r_{\min} = \sqrt{\frac{-2\Delta}{C\,(a + b + R)}},$$

$$\theta = \frac{1}{2}\arctan\left(\frac{2h}{a - b}\right),$$

where

$$R^2 = (a - b)^2 + 4h^2, \qquad C = ab - h^2,$$

$$\Delta = \det\left(\begin{vmatrix} a & h & g \\ h & b & f \\ g & f & 1 \end{vmatrix}\right).$$

c). Objective Function. Optimization refers to deciding on the high-quality element from one set of to be had options. In the handiest case, it approach to minimize an objective characteristic or errors by systematically deciding on the values of variables from their legitimate stages. In order to calculate the mistake produced by way of a candidate answer E, the ellipse coordinates are calculated as a digital form which, in turn, need to additionally be demonstrated, if it really exists in the edge image. The test set is represented by $S = \{s1, s2, \dots s_N\}$ where Ns are the number of points over which the existence of an edge point, corresponding to E, should be tested.

The set S is generated by the Midpoint Ellipse Algorithm (MEA) which is a searching method that seeks required points for drawing an ellipse. For any point (x, y) lying on the boundary of the ellipse with a, h, b, g, and f, it does satisfy the equation

$$f_{eclipse}(x,y) \cong r_{max}x^2 + r_{min}y^2 - r^2{}_{max}r^2{}_{min}$$

where $r$max and $r$min represent the major and minor axes, respectively. However, MEA avoids computing square-root calculations by comparing the pixel separation distances. A method for direct distance comparison is to test the halfway position between two pixels to determine if this midpoint is inside or outside the ellipse boundary. If the point is in the interior of the ellipse, the ellipse function is negative. Thus, if the point is outside the ellipse, the ellipse function is positive.

ISSN No:-2456 –2165

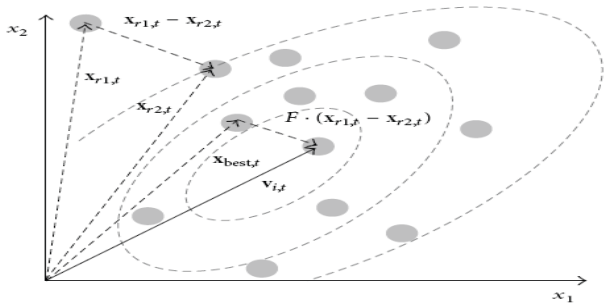


Fig 1: Two-Dimensional Example

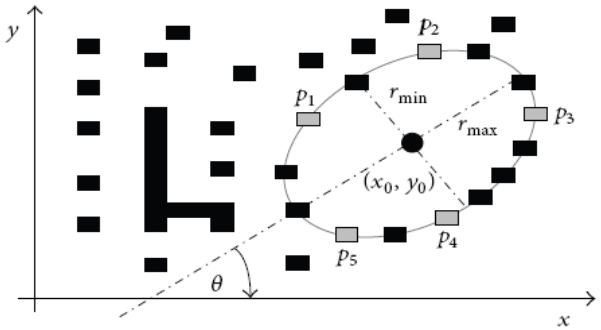


Fig 2: Ellipse Candidate (individual) Built from the Combination of point's $p1$, $p2$, $p3$, p4, and $p5$.

Therefore, the error involved in locating pixel positions using the midpoint test is limited to one-half the pixel separation. To summarize, the relative position of any point (x, y) can be determined by checking the sign of the ellipse function:

$$f_{\text{ellipse}}(x,y)\begin{cases}<0 & \text{if } (x,y) \text{ is inside the ellipse boundary}\\ =0 & \text{if } (x,y) \text{ is on the ellipse boundary}\\ >0 & \text{if } (x,y) \text{ is outside the ellipse boundary.}\end{cases}$$

The ellipse-function test in is applied to mid positions between pixels nearby the ellipse path at each sampling step. The ellipse is used to divide the quadrants into two regions; the limit of the two regions is the point at which the curve has a slope of −1. In Midpoint Ellipse Algorithm the computation time is reduced by considering the symmetry of ellipses. Ellipses sections in adjacent octants within one quadrant are symmetric with respect to the $dy/dy = -1$ line dividing the two octants. These symmetry conditions are illustrated in Figure 4. The algorithm can be considered as the quickest providing a subpixel precision. However, in order to protect the Midpoint Ellipse Algorithm operation, it is important to assure that points lying outside the image plane must not be considered in S. The objective function $J(E)$ represents the matching error produced between the pixels S of the ellipse candidate E and the pixels that actually exist in the edge image, yielding

$$J(E) = 1 - \frac{\sum_{v=1}^{N_s} G(x_v, y_v)}{N_s},$$

Where G(xi,yi) is a function that verifies the pixel existence in x_v, y_v, with $x_v, y_v \in S$ *and* N_s being the number of pixels lying on the perimeter corresponding to E currently under testing. Hence, function G (x_v, y_v) is defined as

$$G(x_v, y_v) = \begin{cases}1 & \text{if the pixel } (x_v, y_v) \text{ is an edge point}\\ 0 & \text{otherwise.}\end{cases}$$

A value of $J(E)$ near to zero implies a better response from the "ellipsoid" operator. The procedure to evaluate a candidate action E with its representation as a virtual shape S. The virtual shape S representing the individual E = {$p1$, p2, p3, p4, p5} the virtual shape S is compared to the original image, point by point, in order to find coincidences between virtual and edge points. The individual has been built from point's $p1$, p2, p3, p4 and p5.

d). Implementation of Deferential Evolution for Ellipse Detection. The ellipse detector algorithm based on Deferential Evolution can be summarized in algorithm1

Algorithm 1

Step 1: Set the DE parameters $F = 0.25$ and CR = 0.8.

Step 2: Initialize the population of m individuals $\mathrm{E}^k = \{E_1^k, E_2^k, \ldots, E_m^k\}$ where each decision variable p_1, p_2, p_3, p_4 and p_5 of E_a^k is set randomly within the interval $[1, N_p]$. All values must be integers. Considering that $k = 0$ and $a \in (1, 2, \ldots, m)$.

Step 3: Evaluate the objective value $J(E_a^k)$ for all m individuals, and determining the $E^{\text{best},k}$ showing the best fitness value, such that $E^{\text{best},k} \in \{\mathrm{E}^k\} \mid J(E^{\text{best},k}) = \min\{J(E_1^k), J(E_2^k), \ldots, J(E_m^k)\}$.

Step 4: Generate the trial population $\mathrm{T} = \{T_1, T_2, \ldots, T_m\}$:

```
for (i = 1; i < m + 1; i++)
do r1 = floor(rand(0, 1) · m); while (r1 = i);
do r2 = floor(rand(0, 1) · m); while ((r2 = i) or (r2 = r2));
jrand = floor(5·rand(0, 1));
        for (j = 1; j < 6; j++) // generate a trial vector
            if (rand(0,1)<=CR or j = jrand)
            T_ji = E_j^(best,k) + F · (E_jr1^k − E_jr2^k);
                else
                    T_ji = E_ji^k;
            end if
        end for
end for
```

Step 5: Evaluate the fitness values $J(T_i)$ $(i \in \{1, 2, \ldots, m\})$ of all trial individuals. Check all individuals. If a candidate parameter set is not physically plausible, i.e. out of the range $[1, N_p]$, then an exaggerated cost function value is returned. This aims to eliminate "unstable" individuals.

Step 6: Select the next population $\mathrm{E}^{k+1} = \{E_1^{k+1}, E_2^{k+1}, \ldots, E_m^{k+1}\}$:

```
for (i = 1; i < m + 1; i++)
    if (J(T_i) < J(E_i^k))
    E_i^(k+1) = T_i
        else
        E_i^(k+1) = E_i^k
    end if
end for
```

Step 7: If the iteration number (N_I) is met, then the output $E^{\text{best},k}$ is the solution (an actual ellipse contained in the image), otherwise go back to Step 3.

B. The Sickle Cell Detector

In order to detect Sickle cell, the proposed detector combines a segmentation strategy with the ellipse detection approach.

a). Image Preprocessing. To employ the proposed detector, smear photos have to be preprocessed to achieve two new photographs: the segmented photograph and its corresponding area map. The segmented photo is produced by way of the usage of a segmentation method while the edge map is generated by using a border extractor algorithm. Such facet map is taken into consideration via the objective function to measure the resemblance of a candidate ellipse with an actual Sickle Cell.

The intention of the segmentation method is to isolate the Sickle cells (Sickle Cell's) from different structures including red blood cells and heritage pixels. Information of color, brightness, and gradients is generally used within a thresholding scheme to generate the labels to classify every pixel.

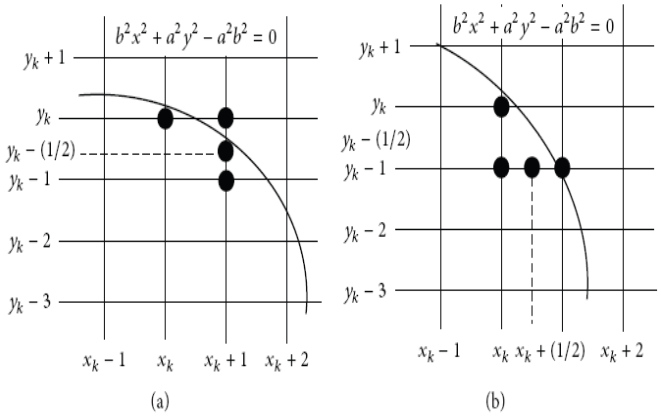


Fig 3: Lableing the Sickle Shape

Although a simple histogram thresholding can be used to segment the SC, at this work the diffused expectation maximization (DEM) has been used to assure better results. DEM is an expectation-maximization- (EM-) based algorithm which has been used to segment complex medical images. In contrast to classical EM algorithms, DEM considers the spatial correlations among pixels as a part of the minimization criteria. Such adaptation allows to segment objects in spite of noisy and complex conditions. The method models an image as a finite mixture, where each mixture component corresponds to a region class and uses a maximum likelihood approach to estimate the parameters for each class, via the expectation maximization (EM) algorithm, which is coupled to anisotropic diffusion over classes in order to account for the spatial dependencies among pixels.

For the SC's segmentation, the implementation of DEM provided in has been used. Since the implementation allows to segment gray-level images and color images, it can be used for operating over all smear images with no regard about how each image has been acquired. The diffused expectation maximization has been configured considering three different classes (K=3) $(\nabla h_{ik} = |\nabla h_{ik}|.^{-9/5}$, and m = 10 iterations.

As a final result of the diffused expectation maximization operation, three different thresholding points are obtained: the first corresponds to the SCs and the second to the red blood cells whereas the third represents the pixels classified as background. The segmentation results obtained by the DEM approach employed at this work considering as the original image. Once the segmented image has been produced, the edge map is computed. The purpose of the edge map is to obtain a simple image representation that preserves object structures.

The Differential Evolution-based detector operates directly over the edge map in order to recognize ellipsoidal shapes. Several algorithms can be used to extract the edge map; however, at this work, the morphological edge detection procedure has been used to accomplish such a task. Morphological edge detection is a traditional method to extract borders from binary images in which original images (I B) are eroded by a simple structure element (I E) composed by a matrix template of 3 × 3 with all its values equal to one. Then, the eroded image is inverted.

(I E) and compared with the original image (I E ∧ I B) in order to detect pixels which are present in both images. Such pixels compose the computed edge map from I B. The edge map obtained by using the morphological edge detection procedure.

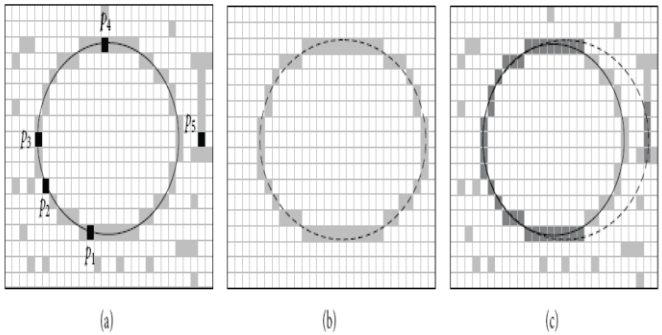
b). Ellipse Detection Approach. The edge map is used as input image for the ellipse detector. The parameter set that has been used in this work for the Differential Evolution algorithm after several calibration examples have been conducted. The final configuration matches the best possible calibration proposed in, where the effect of modifying The Differential Evolution - parameters for several generic optimization problems has been analyzed. The population-size parameter (m =20) has been selected considering the best possible balance between convergence and computational overload. Once it has been set, such configuration has been kept for all test images employed in the experimental study. Under such assumptions, the complete process to detect Sickle Cells is implemented as in Algorithm 2.
Algorithm 2

Step 1: Segment the Sickle Cell using the diffused expectation maximization (DEM) algorithm.
Step 2: Get the edge map from the segmented image.
Step 3: Start the ellipse detector based in Differential Evolution over the edge map while saving best ellipse.
Step 4: Define parameter values for each ellipse that identify the SC.

c). Numerical Example. In order to present the algorithm's step-by-step operation, a numerical example has been set by applying the proposed method to detect a single leukocyte lying inside of a simple image. The image used in the example. After applying the threshold operation, the Sickle

ISSN No:-2456 –2165

Cell is located beside few other pixels which are merely noise. Then, the edge map is subsequently computed and stored pixel by pixel inside the vector *P*.



The Differential Evolution-based ellipse detector is executed using information of the edge map (for the sake of easiness, it only considers a population of four particles). Like all evolutionary approaches, Differential Evolution is a population-based optimizer that attacks the starting point problem by sampling the search space at multiple, randomly chosen and initial particles. By taking five random pixels from vector *P*, four different particles are constructed.

2. Thalassemia Detection

Methodology that will be used in this paper includes:

A. Image Acquisition

Blood image from slides will be obtained using digital microscope.

B. Image Pre-Processing

The main image processing tasks consists of enhancing the image's qualities and deleting overlapped blood cells in the borderline area of the image. Both tasks can be subdivided into smaller tasks.

- Green Plane Extraction: The green plane is extracted from the imported blood cell image. The other planes such as red and blue are not considered because they contain less information about the image.
- Histogram equalization: This process adjusts intensity values of the image by performing histogram equalization involving intensity transformation.
- Contrast and brightness adjustment: To adjust brightness of an image, and histogram of the interested image is used to determine data and display ranges of the image.

C. Image Segmentation

Image segmentation such as threshold-based, edge-based, region-based or clustering methods, such as, fuzzy-C mean clustering and K-mean clustering. Threshold techniques cannot always produce meaningful results since no spatial information is used during the selection of the segmentation threshold [2]. They are often combined with mathematical morphology operations. A simple thresholding approach is applied to give initial labels to pixels in the blood cell images. Then the labels are adjusted with a shape detection method based on large regional context information to produce meaningful results.

The proposed method for the segmentation of blood cell is given below [2].

Step 1: Input the color blood slide image to the system.
Step 2: Convert the color image into gray scale image.
Step 3: Enhance contrast of the gray scale image by histogram equalization method (A).
Step 4: To adjust image intensity level apply linear contrast stretching to gray scale image (B).
Step 5: Obtain the image I1=B+A to brighten all other image components except cell nucleus.
Step 6: Obtain the image I2=I1-A to highlight the entire image objects along with cell nucleus.
Step 7: Obtain the image I3=I1+I2 to remove all other components of blood with minimum effect of Distortion over nucleus. Step 8: To reduce noise, preserve edges and increase the darkness of the nuclei implement 3-by-3 minimum filter on the image I3.
Step 9: Apply a global threshold Otsu's method on image I3.
Step 10: Using the threshold value in above step convert I3 to binary image.
Step 11: To remove small pixel groups use morphological opening.
Step 12: To form objects connect the neighboring pixels.
Step 13: By applying the size test removal of all objects that are less than 50% of average RBC area is done. It is observed that this method of segmentation yields better results than that of previous methods.

D. Feature Extraction

Classification is the task of assigning to the unknown test vector to a known class.
Area: The area was determined by counting the total number of none zero pixels within the image region. Perimeter: It was measured by calculating distance between successive boundary pixels.
Circularity: This is a dimensionless parameter which changes with surface irregularities and is defined as, Circularity = 4* Pi* Area/ Perimerer2

E. Image Classification

Based on the features extracted in above step, classifier classifies the thalassemia cells. Classification is the task of assigning to the unknown test vector, a label from one of the known classes [2]. The k-Nearest- Neighbors (kNN) is a non-parametric method of classification. It is simple but very effective in many cases. Here also kNN has been utilized to classify blast cells from normal white blood cells.

ISSN No:-2456 –2165

IV. RESULTS AND TABLES

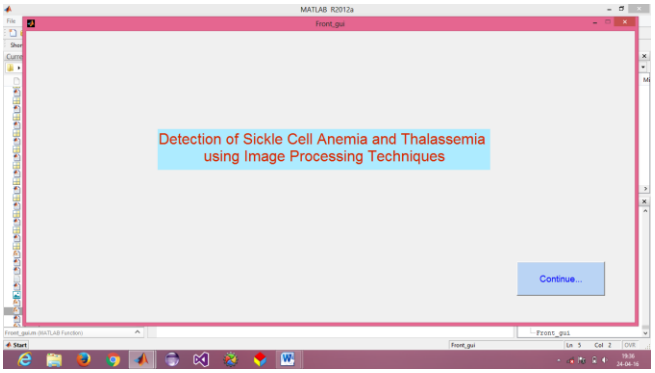


Fig 4: Sickle Cell and Thalassemia Detection Front Sheet.

Front sheet for Sickle Cell and Thalassemia Detection using Image Processing

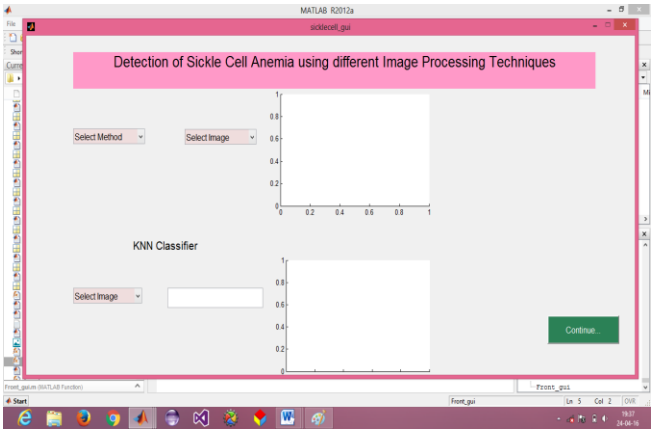


Fig 5: Sickle Cell Detection

The second page contains Sickle Cell Detection. Where we select testing samples and we apply different image processing techniques.

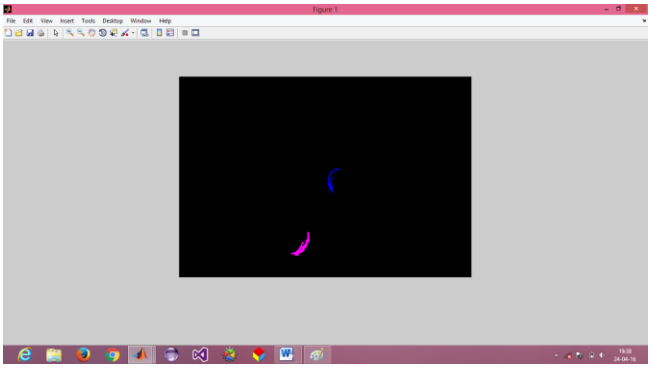


Fig 6: Ellipse Shape Detected Using Ellipse Detector

In order to detect ellipse shapes, candidate images must be preprocessed first by an edge detection algorithm which yields an edge map image. then for ellipse shape detection ellipse detector.

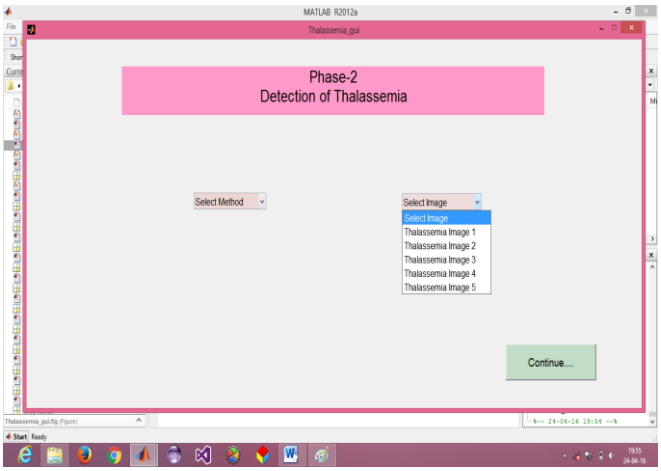


Fig 7: Thalassemia Detection

For Thalassemia Detection for the image segmentation we use fuzzy C and K-mean clustering. Then we take Circularity as a parameter to extract image we use KNN Classifier for classifying Thalassemia Detection.

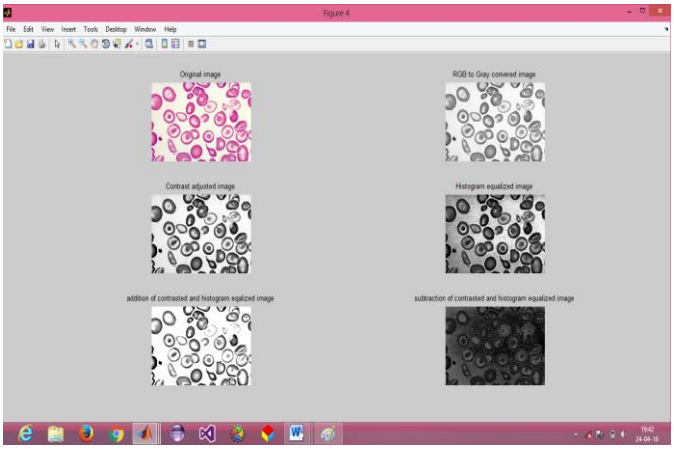


Fig 8: Thalassemia Detection Using Otsu

For detection of Thalassemia we use Otsu method.

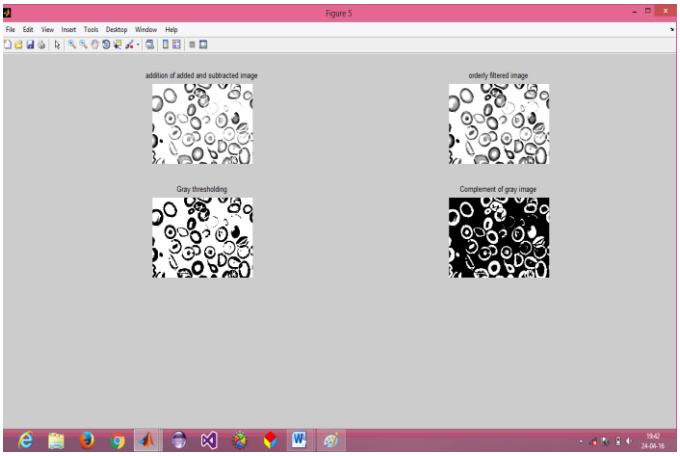


Fig 9: Thalassemia Detection Final Resul

V. CONCLUSION

The proposed method involves detecting the types of RBC disorders using microscopic blood sample images. The system will be built by using features in microscopic images by examining changes on texture, geometry, colors and statistical analysis as a classifier input. Sickle cell disease is detected using DE algorithm. The approach considers the complete process as uses the encoding of five edge points as candidate ellipses in the edge map of the smear. An objective function allows to accurately measuring the resemblance of a candidate ellipse with an actual SC on the image. Guided by the values of such objective function, the set of encoded candidate ellipses are evolved using the DE algorithm so that they can fit into actual SC on the image. The approach generates a sub pixel detector which can effectively identify sickle cell in real images. Thalassemia can be detected using Otsu thresholding. Both methods give a better result compared to previous works.

REFERENCES

[1]. An Improved Computer Vision Method for White Blood Cells Detection. Erik Cuevas,Margarita Díaz,MiguelManzanares, Daniel Zaldivar, andMarco Perez-Cisneros Departament of Electronics.
[2]. Analysis of blood samples for counting thalassemia cells using Support vector machine and nearest neighbor.
[3]. Aruna N.S. , Hariharan S. “Edge Detection of Sickle Cells in Red Blood Cells”,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014,Pages 4140-4144, ISSN 0975-9646.
[4]. J. Poomcokrak and C. Neatpisarnvanit, “Red Blood Cells Extraction and Counting”,The 3rd International Symposium on Biomedical Engineering (ISBME 2008).
[5]. Nasrul Humaimi Mahmood and Muhammad Asraf Mansor, “Red Blood Cells Estimation Using Hough Transform Technique”, Signal & Image Processing: An International Journal (SIPIJ) Vol.3, No.2,April 2012.
[6]. Automated Diagnosis of Thalassemia Based on DataMining Classifiers 1 Eyad H. Elshami, 2 Alaa M. Alhalees Wongseree, W., Chaiyaratana, N., Vichittumaros, K.,Winichagoon, P., Fucharoen, S.:
[7]. Thalassemia classification by neural networks and genetic programming. Information Sciences, vol. 177, pp.771–786,(2007).