# Detection of Anaemic Patients from Microscopic Blood Images with Emphasis on Sickle Cell Anaemia


**Chayashree Patgiri**

Department of Electrical Engineering

Assam Engineering College

This thesis is submitted to

Gauhati University as requirement for the degree of

*Doctor of Philosophy*

Faculty of Engineering

September 2023

I would like to dedicate this thesis to my parents and parents-in-law for continuously being with me as the sources of inspiration and support …

# Declaration

I hereby declare that this thesis is the result of my own research work which has been carried out under the guidance of Prof (Dr.) Amrita Ganguly of Assam Engineering College. I further declare that this thesis as a whole or any part thereof has not been submitted to any university (or institute) for the award of any degree or diploma.

This thesis contains less than 90,000 (ninety thousand) words excluding bibliography and captions.

Chayashree Patgiri

September 2023

# Certificate

This is to certify that the thesis titled "Detection of Anaemic Patients from Microscopic Blood Images with Emphasis on Sickle Cell Anaemia" is the result of research work of Chayashree Patgiri, carried under my supervision, submitted to Gauhati University for the award of the degree of Doctor of Philosophy in Electrical Engineering.

This thesis conforms to the standard of PhD Thesis under Gauhati University including the standard related to plagiarism and has a similarity index not more than 10% (ten percent), excluding the bibiliography.

Prof (Dr.) Amrita Ganguly, Supervisor

September 2023

**Members of the Research Advisory Committee:**

Prof (Dr.) Durlav Hazarika, Department of Electrical Engineering, AEC

Prof (Dr.) Damodar Agarwal, Department of Electrical Engineering, AEC

Prof (Dr.) Bani Kanta Talukdar, Department of Electrical Engineering, AEC

# Acknowledgements

While carrying out the research work of my PhD programme, some persons have provided significant help and support which were the source of my motivation.

At first, I would like to express my gratitude from the deepest corner of my heart to my supervisor Dr. Amrita Ganguly, for keeping faith on me and for giving me the opportunity to carry out research work under her supervision. She has been encouraging me continuously during this period and provided all the necessary supports. Her expertise in the areas of interest helped me to maintain my moral high throughout the journey of this PhD work.

I am thankful to the authorities of Assam Engineering College as well as Gauhati University for providing me all kinds of logistics and administrative support during the programme. I would like to express my sincere thankfulness to all the members of my Research Advisory Committee for continuously reviewing my progress during this period and for giving my research the correct direction. I would like to convey my heartfelt thanks to the faculty members and support staff of the department of Electrical Engineering of Assam Engineering College for their tremendous supports. I am also thankful to my fellow friends who provided me valuable inputs and suggestions at every moment whenever required.

At this juncture of my PhD work, I want to express my gratitude towards my family for encouraging me to achieve this goal of my life through their blessings, love and affection. I would also like to express my honest and heartfelt thankfulness towards my husband for being my inspiration and strength throughout this journey and helping me to fulfil the aspirations of my family.

Finally I would like to thank the almighty for showing me the correct path at every step of my research work and for providing me good health.

Chayashree Patgiri

September 2023

# Abstract

Anaemia is a blood disorder with reduced numbers of healthy red blood cells and therefore lowering oxygen carrying capacity of blood. It is one of the most common blood disease being experienced by millions of people worldwide. Sickle cell anaemia (SCA) is a type of anaemia where human body produces defective or deformed red blood cells and thereby reducing the amount of oxygen transmitting through the blood for proper functioning of the body. It is an inherited type of disease in which blood protein i.e. haemoglobin becomes defective. These abnormal haemoglobins are passed from parents to their children. The abnormal haemoglobins are called haemoglobin S or sickle haemoglobin. These are caused by mutations in the $\beta$-globin gene which results in the change in shape of the red blood cells. The faulty red blood cells become crescent shaped, sticky and rigid instead of round and flexible thus creating problem in flowing through the blood vessels having specific diameters in human body. SCA has become a global health burden with millions of people affected by this blood disorder every year. Mostly it is predominant among the people who are from or whose ancestors are from regions like Central and South America, the Caribbean, Saudi Arabia, India, Mediterranean countries such as Greece, Italy, Turkey and sub-Saharan Africa. In India, primarily the tribal populated regions are endemic to sickle cell disease with prevalence rate ranging from 5% to 40%. The states of India such as Gujarat, Madhya Pradesh, Tamil Nadu, Kerala, West Bengal, Odisha, Assam, Maharashtra and Chhattisgarh are mostly affected by this disease.

In view of this, there is certain need for designing comprehensive health programs to tackle sickle cell disease. Effective diagnosis techniques also play a crucial role in attaining

lower prevalence of the disease. As sickle cell anaemia is an incurable inherited disorder, early detection of the disease is very much important in order to prevent its complicated progressive health issues. Proper treatment can be provided to the patients only if the disease can be diagnosed at an early stage. It is also evident that the mortality of children under the age of 5 years can be significantly decreased if proper screening of newborns is done at their birth time. In many parts of the world, most of the people born with SCD die before diagnosis in early childhood. Digital image processing techniques play a vital role in diagnosis of disease in medical field. These techniques can be used in automatic detection process of this disease from microscopic blood images of patients. This replaces tedious, time consuming manual inspection of abnormal cells. From early detection point of view, the automatic, robust detection of sickle cells using various image processing techniques are proved to be very much effective.

The proposed work here focuses on automatic detection of SCA from microscopic blood images using image processing techniques. Firstly, for segmentation purpose adaptive thresholding techniques are experimented with the input images. Four different types of local adaptive thresholding methods are applied and results are compared with Otsu's global thresholding method. The different adaptive thresholding methods used here are namely Bernsen, Sauvola, Niblack and NICK thresholding. These adaptive thresholding methods are clubbed with fuzzy C-means clustering and active contour-based technique separately in order to form two varieties of hybrid segmentation methodologies for effective segmentation. Thereafter, features were extracted from the segmented images of blood. In this work, eight different shapes or geometrical features namely area, perimeter, form factor (metric value), minor axis, major axis, aspect ratio, eccentricity and solidity are selected and extracted. These features are used to form the feature vector for classification purpose. After extracting the geometrical features of cells from segmented images, different supervised machine learning methods are used for classification of normal and abnormal sickle RBCs.

Two supervised learning methods, Naïve Bayes and K-nearest neighbor (KNN) classifiers are implemented with first variety of hybrid segmentation methodology comprising adaptive thresholding and fuzzy C-means clustering. The results are compared for different threshold-

ing methods using the performance parameters accuracy, sensitivity, specificity and precision. Out of four adaptive thresholding techniques, NICK's method gives better performance in case of both Naïve Bayes and KNN classifiers. It is also seen that, the performance of KNN classifier is comparatively superior to that of the Naïve Bayes classifier. The KNN classifier gives the best performance in terms of the considered performance parameters with hybrid segmentation method combining fuzzy C-means and NICK's thresholding. This combination yields a significant classification accuracy of 98.87% with lowest computational time.

For improving the performance of the system further, second variety of hybrid segmentation methodology comprising adaptive thresholding and active contour-based methods are implemented with two other supervised learning methods namely artificial neural network (ANN) and support vector machine (SVM) classifiers. This approach achieves a maximum of 99.2% accuracy in case of SVM classifier with NICK's adaptive thresholding and active contour-based hybrid segmentation method. The performance of ANN classifier is studied for seven different training algorithms by varying the numbers of hidden layer neurons. Out of the seven algorithms, resilient back-propagation algorithm with 10 numbers of hidden neurons gave better performance in ANN with 99% accuracy rate. A comparative analysis is also carried out for ANN and SVM classifier with Otsu's global thresholding and active contour-based hybrid segmentation method. The performance of ANN and SVM classifiers with adaptive thresholding method is found to be better than that of the global thresholding method.

Thirdly, extreme learning machine (ELM) classifier is used with hybrid segmentation comprising four adaptive thresholding and active contour-based methods. The system gives 99.4% overall accuracy and computation time of 0.15 second using active contour-based method in combination with NICK's thresholding which makes it a fast, robust and effective SCA detection system.

The use of different image processing techniques replaces the time consuming manual process of counting and detection of abnormal cells by the laboratory technician, which may results in erroneous results. The modern digital image processing methods give accurate results with faster computational time. In this work, we have presented applications of

different supervised machine learning techniques for efficient classification of SCA. Also hybrid segmentation framework comprising different adaptive thresholding techniques along with clustering and region based methods are applied on the image database. The results evaluated and analysed in this work clearly show that these supervised learning algorithms with hybrid segmentation techniques are very effective means for classification of SCA.

# Table of contents

**List of figures** **xix**

**List of tables** **xxiii**

**1 Introduction** **1**

1.1 Background . . . . . 1

1.2 Literature survey . . . . . 6

1.3 Motivation . . . . . 14

1.4 Problem Statement . . . . . 16

1.5 Contribution . . . . . 17

1.6 Organization of the thesis . . . . . 18

**2 Theoretical Background** **19**

2.1 Sickle cell anaemia . . . . . 20

2.2 Image segmentation . . . . . 24

2.2.1 Thresholding based segmentation . . . . . 25

2.2.2 Clustering based segmentation . . . . . 29

2.2.3 Region based segmentation . . . . . 32

2.3 Supervised machine learning . . . . . 33

2.3.1 Naïve Bayes classifier . . . . . 33
2.3.2 K-nearest neighbor classifier . . . . . 35
2.3.3 Artificial neural network . . . . . 35
2.3.4 Support vector machine . . . . . 37
2.3.5 Extreme learning machine . . . . . 37

**3 Detection of Red Blood Cell and Sickle Cell from Microscopic Blood Images using Local Adaptive Thresholding Techniques 41**

3.1 Introduction . . . . . 41
3.2 Detection of sickle cells using Niblack's adaptive thresholding method . . . 42
3.2.1 System Block Diagram . . . . . 42
3.2.2 Results and Discussion . . . . . 50
3.3 Comparative study on different adaptive thresholding techniques for detection of sickle cells . . . . . 52
3.3.1 System Block Diagram . . . . . 52
3.3.2 Results and Discussion . . . . . 55

**4 Classification of Red Blood Cell and Sickle Cell Using Naïve Bayes Classifier and K-Nearest Neighbor Classifier 59**

4.1 Introduction . . . . . 59
4.2 System Block Diagram . . . . . 61
4.2.1 Image pre-processing . . . . . 61
4.2.2 Image Segmentation . . . . . 63
4.2.3 Extraction of features from the segmented images . . . . . 64
4.2.4 Classification using Naïve Bayes and KNN classifiers . . . . . 65
4.3 Results and Discussion . . . . . 68

**5 Classification of Sickle Cell based on Adaptive Thresholding and Contour-based methods using ANN and SVM Classifiers 81**

5.1 Introduction . . . . . 82

5.2 System Block Diagram . . . . . 83

5.2.1 Pre-processing of input data sample . . . . . 84

5.2.2 Segmentation of pre-processed images . . . . . 85

5.2.3 Feature extraction from segmented images . . . . . 86

5.2.4 Classification of images using ANN and SVM classifiers . . . . . 90

5.3 Results and Discussion . . . . . 90

**6 Sickle Cell Detection from Blood Images of Anaemic Person using Hybrid Segmentation Method and ELM Classifier 101**

6.1 Introduction . . . . . 101

6.2 System Block Diagram . . . . . 103

6.2.1 Image segmentation using hybrid method . . . . . 104

6.2.2 Feature extraction from segmented images . . . . . 105

6.2.3 Classification using ELM classifier . . . . . 105

6.3 Results and Discussion . . . . . 108

**7 Conclusion 113**

7.1 Conclusion . . . . . 113

7.2 Future scope . . . . . 117

**References 119**

**Appendix A List of Publications 133**

# List of figures


1.1 Spread of sickle cell anaemia worldwide (Source: F. B. Piel *et al.* Sickle cell disease, The New England Journal of Medicine **376**, 1561-1573, 2017) . . . 15

1.2 Prevalence of Sickle cell disease in India (Source: Global Burden of Disease (GBD) "Invisible" Pain endured by millions of patients across India) . . . . 16

2.1 Normal and sickle red blood cell . . . . . 21

2.2 Mutation in sickle cell disease . . . . . 23

2.3 Types of sickle cell disease . . . . . 25

2.4 Sickle cell inheritance . . . . . 26

2.5 Segmentation methods for sickle cell anaemia . . . . . 27

2.6 (a) Original image (b) Histogram of image (c) Result of global thresholding 28

2.7 (a) Original image (b) Result of global thresholding (c) Image subdivided into individual sub-images (d) Result of adaptive thresholding . . . . . 30

2.8 Example of clustering . . . . . 30

2.9 Different layers of ANN . . . . . 36

2.10 SVM . . . . . 37

2.11 ELM structure. . . . . 38

3.1 Block diagram for detection of sickle cells . . . . . 43
3.2 Sample microscopic blood images of sickle cell anaemic patients . . . . . 45
3.3 Sample microscopic blood images of sickle cell anaemic patients . . . . . 46
3.4 (a) Sample image (b) Segmented image using Niblack's thresholding (c) Final segmented image . . . . . 49
3.5 (a) Image 1 (b) Detected normal RBCs and sickle cells of image 1 (c) Image 2 (d) Detected normal RBCs and sickle cells of image 2 (e) Image 3 (f) Detected normal RBCs and sickle cells of image 3 . . . . . 51
3.6 System diagram for detection of sickle cells . . . . . 53
3.7 (a) Sample image (b) Segmented image using Bernsen's thresholding (c) Segmented image using Niblack's thresholding (d) Segmented image using Sauvola's thresholding (e) Segmented image using NICK's thresholding . . 56

4.1 System block diagram of the proposed system . . . . . 62
4.2 (a) Sample image (b) Result of hybrid segmentation using Bernsen's thresholding (c) Result of hybrid segmentation using Niblack's thresholding (d) Result of hybrid segmentation using Sauvola's thresholding (e) Result of hybrid segmentation using NICK's thresholding . . . . . 69
4.3 Confusion matrix for calculation of performance parameters . . . . . 70
4.4 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using four thresholding methods . . . . . 77

5.1 System block diagram . . . . . 84
5.2 (a) Sample image (b) Result of hybrid segmentation using Otsu's global thresholding (c) Result of hybrid segmentation using Bernsen's thresholding (d) Result of hybrid segmentation using Niblack's thresholding (e) Result of hybrid segmentation using Sauvola's thresholding (f) Result of hybrid segmentation using NICK's thresholding . . . . . 88

5.3 Extracted features of normal and sickle RBCs from a cropped sample image [Blue colour=Major axis, Yellow colour=Minor axis, , White color=Aspect Ratio, Magenta colour=Solidity, Black colour=Area, Red colour=Perimeter, Green color=Metric value , Cyan color=Eccentricity] . . . . . 89
5.4 Confusion matrix of ANN classifier . . . . . 92

6.1 System block diagram for normal RBC and sickle cell classification using ELM 103
6.2 Overall performance of ELM classifier . . . . . 112

# List of tables


3.1 Accuracy of the thresholding techniques . . . . . 57

3.2 Computation time of different thresholding techniques . . . . . 57

4.1 Extracted features of a sample image using hybrid segmentation method . . 67

4.2 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Bernsen thresholding . . . . . 72

4.3 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Sauvola thresholding . . . . . 73

4.4 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Niblack thresholding . . . . . 74

4.5 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using NICK thresholding . . . . . 75

4.6 Overall performance of Naïve Bayes and KNN classifier with hybrid segmentation using four thresholding methods . . . . . 76

4.7 Computation time of Naïve Bayes and KNN classifier . . . . . 78

4.8 Comparison of performance of our method with other existing methods . . 79

5.1 Feature values for a normal and sickle cell . . . . . 90

5.2 Performance of ANN and SVM classifier using Bernsen thresholding and active contour-based segmentation . . . . . 93
5.3 Performance of ANN and SVM classifier using Sauvola thresholding and active contour-based segmentation . . . . . 94
5.4 Performance of ANN and SVM classifier using Niblack thresholding and active contour-based segmentation . . . . . 95
5.5 Performance of ANN and SVM classifier using NICK's thresholding and active contour-based segmentation . . . . . 96
5.6 Overall performance of ANN and SVM classifier using four adaptive thresholding methods and active contour-based segmentation . . . . . 97
5.7 Overall performance of ANN and SVM classifier using Otsu's thresholding and active contour-based segmentation . . . . . 98
5.8 Overall Accuracy of ANN classifier with different training algorithms and varying hidden layer neurons . . . . . 99
5.9 Performance comparison of proposed system with other techniques . . . . . 99

6.1 Shape features of sample normal and sickle RBCs . . . . . 107
6.2 Performance of ELM classifier with hybrid segmentation using Bernsen thresholding . . . . . 108
6.3 Performance of ELM classifier with hybrid segmentation using Sauvola thresholding . . . . . 109
6.4 Performance of ELM classifier with hybrid segmentation using Niblack thresholding . . . . . 110
6.5 Performance of ELM classifier with hybrid segmentation using NICK thresholding . . . . . 111
6.6 Overall performance of ELM classifier using hybrid segmentation . . . . . 111
6.7 Computation time of ELM classifier with different thresholding method . . 111

7.1 Comparative performance analysis of different methods considered for the detection of SCA . . . . . 116

# 1

# Introduction

This chapter gives a brief overview of related background of sickle cell anaemia, literature survey, motivation, problem formulation, contribution and organization of the thesis.

## 1.1 Background

Anaemia is a blood disorder with reduce number of healthy red blood cells and therefore lowering the capacity of transmitting oxygen through blood. It is one of most common blood disease worldwide causing millions of people. Sickle cell anaemia is a type of anaemia where human body produces defective or deformed red blood cells and thereby reducing the

amount of oxygen for proper functioning of the body. It is an inherited type of disease in which blood protein i.e. haemoglobin becomes defective. These abnormal haemoglobins are passed from parents to their children. The abnormal haemoglobin is called haemoglobin S or sickle haemoglobin caused by mutations in the $\beta$-globin gene which results in the change in shape of red blood cells. The faulty red blood cells become crescent shaped, sticky and rigid instead of round and flexible thus creating problem in flowing through the blood vessels in human body. The organs of body do not get enough oxygen because of sickling of blood cells that can cause organ damage. Episodes of pain in different organs, fatigue, yellowing of skin, shortness of breath, infections, swelling of the hands and feet are major symptoms of sickle cell anaemia. The life spans of normal red blood cells are 90 to 120 days, whereas sickle cells die in 10 to 20 days, causing scarcity of red blood cells. A person with sickle cell disease inherits two haemoglobin S genes, one from each parent. But if a person inherits one defective haemoglobin gene from one parent and one normal haemoglobin gene from other parent, then the person is called having sickle cell trait. These types of persons do not show any serious health complications, living normal life but they are carriers of defective gene, so they can pass the abnormal haemoglobin gene to their children. Various complications arise due to sickle cell anaemia including brain stroke, pulmonary hypertension, vision problem, retarded growth in children, hand and foot disease, severe anaemia, heart disease etc.

Sickle cell disease affects millions of individuals worldwide. Mostly it is predominant among people who come from or whose ancestors come from regions like central and south America, the Caribbean, Saudi Arabia, India, Mediterranean countries such as Greece, Italy, Turkey and sub-Saharan Africa [1, 2]. According to World Health Organization (WHO), haemoglobin disorder affects around 5% of world's population [3]. Among haemoglobin disorder, sickle cell anemia and thalassemia are most commonly widespread in the affected regions of the world. Approximately 3,00,000 new born babies are affected with severe inherited blood disease mainly sickle cell anaemia in every year globally and 80% of them are

born in middle and low income countries. African countries like Nigeria, Gabon, Cameroon, Ghana, Republic of Congo and Uganda has highest rate of this disease with $20\% - 30\%$ [4]. In central and south America, African origins Americans are more susceptible to be infected with this disease. The disease becomes a serious health burden for countries with high prevalence rate. According to Centers for disease control and prevention, the disease is estimated to affect 1 out of 365 African origin Americans people and 1 in 13 African origin Americans are carriers of sickle cell traits [5]. In [6], E. Wastnedge et al. presents a detailed study of prevalence of sickle cell disease in children under 5 years of age. They found that Africa has highest numbers of SCD people and highest mortality of SCD with 1125.49 per 1,00,000 live births. They concluded their review with importance of requiring newborn screening programmes to tackle the burden of SCD in early childhood. They also give importance of increasing the study and research on this haematological disorder and proper addressing of this under recognised disease by the international bodies. In study [7], authors reported the increasing number of SCD patients in middle or low income countries like India, Nigeria, and the Democratic Republic of the Congo because of poor infrastructure of public health facilities and proper screening programmes. They highlighted the necessity of awareness programmes on SCD, improved public health infrastructure and adequate funding for comprehensive programmes on screening of newborns. It is approximated that, almost 50% of sickle cell trait and sickle cell anaemic newborn babies are from Nigeria, Democratic Republic of the Congo and India. Among these, about 15% of world's sickle cell diseased newborn babies are from India [8].

In India, mainly tribal populated regions are endemic to sickle cell disease with ranging from 5% to 40% prevalence rate [9]. The states of India such as Gujarat, Madhya Pradesh, Tamil Nadu, Kerala, West Bengal, Odisha, Assam, Maharashtra and Chhattisgarh are mostly affected by the disease [10]. In scheduled tribe (ST) population of western, central and southern India reported 1 in 86 newborn with sickle cell disease. Ministry of Tribal Af-

fairs and Indian Council of Medical Research conducted testing in India and found out 8.75% positive cases of the disease in year 2016-2018. In [11], authors focus on impact of SCD on tribal communities of India for urgent implementation of SCD programmes at national level. Different states in India like Gujarat, Maharashtra, Chhattisgarh, Tamil Nadu, Madhya Pradesh are focusing on screening of population to detect sickle cell disease for providing counselling and treatment at early age of the patient [12]. Under National Rural Health Mission, government of India has issued guidelines for precaution and control of haemoglobinopathies and provides scheme for early detection of the disease. But this is not sufficient for lower-middle income country like India as haemoglobinopathies become major genetic health burden now-a-days. As the disease is spreaded in tribal and non-tribal ethnic groups of all over India, many states require extensive programmes on educating people on SCD and comprehensive programme for diagnosis and treatment to eradicate the disease.

Sickle cell disease is incurable. But with proper diagnosis and effective treatment, the complications of the disease can be reduced and life expectancy of persons suffering from this disease can be increased. For accurate preventative treatment of this life-threatening disease, early diagnosis is very much important. As it is an inherited disease, newborn babies carry defective gene at birth. So, there should be early screening or diagnosis of newborn babies as symptoms of the disease are showing after 5 or 6 months of age, thereby avoiding serious health problems of the disease. This disease can be tested using haemoglobin electrophoresis which measures levels of haemoglobin and examine the types of haemoglobin. The disease becomes severe over the time if not treated in the early age of the patient. Different treatment available depending upon the severity of the patient that can prevent complications and increase the life span of those suffering from the disease. Hydroxyurea, L-glutamine oral powder, voxelotor, crizanlizumab are some of the medicine approved by Food and Drug Administration (FDA) for adult people and children whose age is more than five. These medicines cannot cure the disease, but can reduce the complications. But stem cell transplant

can actually cure sickle cell anemia in which bone marrow of sickle cell anaemia patient is replace with bone marrow from a healthy matched donor. This procedure is very complicated, costly and risky also if the body of the patient rejects transplant that can cause life-threatening sickness or even death [13–15].

Image processing techniques plays a vital role in diagnosis of disease in medical field. It is a non-invasive procedure for disease detection. Medical imaging is a field that is growing rapidly because of advancements in image processing techniques including enhancement, segmentation, representation, analysis, restoration and recognition. These techniques become tremendously beneficial for automatic, accurate and faster diagnosis and treatment of various diseases over the time. In sickle cell anaemia also these image processing techniques are developed as very useful for identification of abnormal red blood cells from the microscopic blood images of sickle cell anemic patient. The work here presents an automatic identification and classification of normal and sickle cells from microscopic blood images using different image processing methods. As this disease needs early detection for proper treatment, the methods proposed in the work can be helpful for physician providing a robust and fast automatic detection process. This work focuses on detection of sickle cells using global and local adaptive thresholding segmentation methods. Four different adaptive thresholding methods are studied for the work and applied on the images. Also two hybrid segmentation methods are also presented in this work. One hybrid segmentation method comprises of fuzzy C-means in combination with adaptive thresholding and another hybrid method comprises active contour-based method and adaptive thresholding method. Different machine learning techniques are applied for automatic classification of normal red blood cells and sickle cells. A comparative analysis of performance of these machine learning methods is carried out.

## 1.2 Literature survey

In this section, a detail review of the literature related to detection of sickle cells is discussed. The literature review is broadly divided into two groups: a) literature related to segmentation of normal RBCs and sickle cells detection and b) literature related to classification of normal and sickle cells from microscopic images.

Many segmentation techniques are proposed and implemented by researchers in the last few years. These are mostly based on thresholding method, edge based method, watershed methods, CHT method, clustering method etc. Some of these are discussed in the following section.

A study based on image processing techniques to detect sickle cells for diagnosis of SCA has been presented in [16]. The authors employed Otsu's thresholding for converting RGB image to binary image and followed by some filters to remove noise. Then the binary image was segmented using CHT and morphological operation to separate sickle cells from the image. They obtained 91.11% accuracy, 92.9% precision and 79.05% recall for the proposed system.

An algorithm for sickle cell segmentation and detection was discussed in [17]. Authors used shape parameters form factor and eccentricity as features for classify normal and abnormal cells using different segmentation methods like Otsu's thresholding, edge detection technique and watershed technique. They obtained Otsu's segmentation gave best results out of methods considered.

In [18], authors have presented a preliminary study about segmentation of red blood cells for counting of cells automatically. They first applied Ycbcr color conversion, morphological operation along with masking to separate white blood cell and then marker-controlled watershed technique for separation of touching cells in the blood smear. Another work presented in [19] also used marker-controlled watershed method for erythrocyte segmentation.

In [20], author described blood segmentation method using median filtering for noise removal, sobel operator for edge enhancement and K-mean clustering for segmentation on stained images. For non stained images, before K-mean clustering they applied thresholding on hue saturation value (HSV) color space converted images. Authors also employed watershed segmentation for clusters of overlapping cells.

In [21], authors proposed an algorithm for identification of borders of cells from blood smear images. They applied Dijkstra's shortest path algorithm with an alternate graph representation which is tested on 98 images and successfully segmented normal, abnormal, touched RBCs, WBCs from blood smear images.

H. Berge et al. described a technique for counting of RBCs from images of blood smear using iterative thresholding with morphological operations for segmentation of RBCs. The Zack's thresholding technique was employed in this work. They calculated curvature of boundary and Delaunay triangulation for separation of RBC clump in the image [22].

In [23], authors have implemented gradient watershed transformation to segment the blood cells from the microscopic images. They obtained 91% accuracy and 92.4% precision for boundary detection of normal and sickle cells.

Fadhel et al. illustrated image processing based techniques to count normal and abnormal RBCs using CHT and watershed segmentation method. They obtained that CHT takes less computation time than watershed method [24].

In paper [25], an automatic RBC counting algorithm replacing the traditional manual counting method is presented by the authors. They used thresholding and Hough transform for the purpose. The paper also demonstrated graphical user interface (GUI) for RBC counting using MATLAB environment.

In [26], an improved watershed segmentation method is proposed for sickle cell detection. They overcome the over segmentation problem of watershed method by applying filtering

of images using de-noising method. They examined accuracy, specificity and sensitivity as performance parameters for the presented work.

In [27], authors proposed hybrid type of segmentation method which is consist of contour-based algorithm and Gram-Schmidt orthogonalization for RBC segmentation from microscopic images. In the image pre-processing step, authors have used Wiener filter and forward discrete curvelet transform for enhancement and noise removal of input image respectively.

In paper [28], authors presented a procedure for investigation of shape of RBC from blood images. They utilized three different types of blood images for validation of the proposed method. They applied ellipse adjustment method for effective identification of round RBCs and crescent shaped cells from clusters of cells.

In [29], authors have presented Sobel edge detection based system for RBC shape determination. They implemented the work in MATLAB environment.

In [30], authors employed Weiner filter for noise removal and Sobel Edge based segmentation method to calculate the metric value of each of the blood cells from microscopic images. They obtained an accuracy of 95.8% for the proposed system.

Different edge detection techniques are proved to be useful in segmentation of red blood cells from microscopic images. A study on application of Laplacian of Gaussian (LoG) edge detection algorithm was done by Mohamad et al. in paper [31] for detection of sickle cell anaemia patients.

In [32], author reported a study on segmentation of RBC for thalassemia disease using unsupervised machine learning method. The employed moving K-mean clustering technique for segmentation and median filtering for removing noises. Further they applied seeded region growing area extraction algorithm to eliminate undesired regions which were not cleaned by the median filter.

In [33], authors have applied different edge detection methods for sickle cells detection. They presented a comparative study of different edge detectors namely Sobel, Prewitt, Canny, Laplacian of Gaussian and Roberts operator.

K. Alotaibi illustrated a method for sickle cells detection from blood smears using image segmentation technique. The author calculated form factor to classify the cell as normal and sickle cell [34].

W. Delgado-Font et al. illustrated a Chan-Vese active contour model for segmentation of region of interest in [35]. They used circular and elliptical shape factor as features for classification of normal, elongated or other deformed RBCs in non-clustered samples. Whereas for clustered cells elliptical adjustment method was employed.

In [36], author focused on automatic segmentation and counting of RBCs and WBCs from microscopic blood images. They applied Otsu, Niblack, Bernsen and Sauvola for separating foreground and background after de-noising and edge enhancement.

In [37], authors proposed a method for classifying and counting RBCs based on geometrical features area and eccentricity of cells. They applied canny edge detector for segmentation. They also compared results with CHT segmentation method and found that the proposed method was superior in terms of counting of cells and computation time.

In [38], three types of noise removal methods were discussed and performances of these methods were compared for finding proper and effective technique for noise reduction from medical images. The noise removal filters used here were ordinary, median and Weiner filter. The performance of median filter was obtained higher than the other two types of filter for sickle cell images.

Various literatures are available associated with classification of red blood cells as normal or abnormal types. There are several machine learning algorithms like artificial neural networks (ANN), support vector machine (SVM), K-nearest neighbor (KNN), Naïve Bayes, convolutional neural network (CNN), deep CNN (DCNN), extreme learning machine (ELM),

random forest, decision tree etc. are very popular among researchers in recent times. The following are some of the recent work on classification of SCA.

A method of feature extraction and categorization of normal and abnormal RBCs applying back propagation neural network was proposed in [39]. A total of 27 features were extracted here including $1^{st}$, $2^{nd}$ order intensity level statistical parameters, 7 moment invariants and 16 geometrical parameters.

In [40], author presented an algorithm for abnormal cells detection like sickle cell and elliptocytosis using circular Hough transform (CHT). For detection process, author applied neural network and obtained an overall accuracy of 98.9% for all types of input cells.

M. Khalaf et al. introduced different recurrent neural network architectures for classifying sickle cells. They implemented Elman and Jordan recurrent neural network and combination of both networks to form a hybrid structure. The structure Jordan recurrent network provides better results compared to other recurrent networks [41].

In [42], author used CHT to segment normal and abnormal cells using shape features like area, perimeter, eccentricity and solidity. He trained and tested neural network and decision tree for classification and regression purpose.

An automatic and robust method for classification of RBC from blood images was proposed in paper [43]. Otsu's global thresholding method is applied followed by noise and holes removal to extract geometrical features from RBCs. Finally authors have applied artificial neural network (ANN) to classify normal and abnormal RBCs which yielded 83% accuracy, 82% precision and 76% recall.

In [44], author used 21 microscopic images for detecting three types of abnormal cells including sickle cells with the help of CHT algorithm and watershed segmentation. Author implements neural network for classification purpose with high accuracy rate.

In paper [45], authors employed various neural network techniques for determining the severity of diseased persons having sickle cells. They used eight different types of

classifiers with two hybrid classifiers comprising Levenberg-Marquardt learning neural network, random forest, Fischer discriminate analysis.

In [46], authors described CNN and RNN combined technique for classification of RBCs. First the CNN layer was pre-trained with image dataset, followed by transfer learning method for initialization of a CNN. Secondly, RNN layer was trained. Finally CNN and RNN generated features were merged.

In [47], authors illustrated a back-propagation based neural network algorithm for pain management of sickle cell disease. The neural network classify four different levels of pain (less/severe acute and chronic pain) based on 15 different attributes.

A work on SCA classification using different types of neural network models were illustrated in [48]. The study used four types of neural network structures: back-propagation feed-forward NN, radial basis NN, functional link NN and voted perceptron classifier. Out of these network structures, they obtained best performance with back-propagation feed-forward NN classifier.

In [49], normal and abnormal RBCs were identified using SVM classifier for making a robust system by the authors. Edge detection method was used for segmentation. They extracted a total of 271 types of features containing texture parameters, geometrical parameters and colour parameter from blood cells.

Chy and Rahaman implemented SVM classifier for automatic classification of deformed cells (sickle cell) in paper [50]. They applied Otsu's thresholding for segmentation and circularity and aspect ratio were used as features to training and testing SVM classifier.

In [51], a decision system was proposed by the authors for classification of different types of anaemia. For extracting WBCs from image authors have used K-Medoids algorithm and granulometric analysis to separate RBCs from WBCs. They have extracted 8 different features of RBCs which help in classification of anaemia.

In [52], author have first applied CHT and watershed segmentation for enhancing and extracting features from 45 microscopic images taken from 15 samples of anaemic persons. The author extracted 10 information data variables by applying cell signature method and these were used to train and test back propagation neural network (BPNN), support vector machine (SVM) and self-organising map (SOM) neural network for detecting normal, sickle, elliptocytosis and burr cells.

Sharma et al. presented K-nearest neighbor (KNN) classifier based detection method for three types of defected RBC cells with 100 images. The features used for the work were form factor, variance of radial signature and aspect ratio of blood cells. They obtained accuracy of 80.6% and sensitivity of 87.2% [53].

X. Gual-Arnau et al. presented an automatic erythrocyte shape detection method with 17 peripheral blood sample and 45 different field images. Four different contour and region based descriptors were extracted from cells for feature vector generation. They applied region growing based segmentation method i.e. active contours and KNN to classify normal, sickle and other deformed cells [54].

In [55], decision tree based method was implemented for determination of sickle cell anemia disease. Image segmentation was done using K-medoids technique with modified watershed transform for separation of WBC nucleus and Otsu's thresholding for RBCs separation.

Lavanya and Sushritha presented an image processing based algorithm for sickle cell anemia and thalassemia detection. They used a novel method called ellipse detection based on deferential evaluation for sickle cell and Otsu's thresholding with KNN classifier for thalassemia detection [56].

In paper [57], authors have demonstration an automated RBC detection process consisting of three stages. First region of interest from RBCs are extracted while separating the overlapping RBCs using automatic seed generation. Secondly they used normalization

technique for RBCs and finally applied deep convolutional neural networks (DCNNs) for classification.

In [58], a deep learning algorithm was discussed for RBC classify into 3 types: normal, sickle and other. The problem of insufficient training data was discussed in the paper and authors applied two method to solve the problem i.e. transfer learning and data augmentation. They obtained a highest accuracy of 99.54%.

Aliyu et al. presented deep learning AlexNet model for identification of 15 different shapes of RBCs including sickle shaped RBCs [59]. They applied Otsu's thresholding and morphological operations in the pre-processing step before feeding images into the AlexNet model to detect normal and abnormal types of RBCs.

An effective method for blood cells segmentation using contour aware neural network was described in paper [60]. The performance of extreme learning machine (ELM) classifier was estimated for $64,000$ blood samples and the proposed technique yielded accuracy of 94.71% and 98.68% for RBS and WBC respectively.

In [61], authors have experimented SVM, KNN and extreme learning machine (ELM) classifiers for detection of sickle cell anemia. Segmentation process was done using fuzzy C-mean clustering along with some morphological operations. They extracted statistical and geometrical features from normal and defected blood cells for training and testing the classifiers.

P. K. Das et al. presented a review work on available literature on image enhancement, segmentation, feature extraction and classification for sickle cell anemia detection. They also discussed the strengths and weaknesses of recent methods used by researchers. The review work also focused on scope for future on sickle cell detection [62].

## 1.3 Motivation

Sickle cell anaemia is a genetic blood disorder which affects the normal red blood cell to change its shape and becomes rigid, crescent shaped cell. These type of deformed RBCs prevent the normal blood circulation through the small veins in the human body. This creates problem of supplying oxygen in the body tissues as sickle shaped red blood cells stuck in the small blood vessels. The disease arises serious health risk such as heart attack, blindness, severe anaemia, stroke, organ damage, infections etc. if not treated in time. This leads to increase in infant mortality rate, low quality of life and short life span. The prevalence rate of this disease is higher among the people of Africa origin American of Central and South America, African countries such as Ghana, Nigeria, Republic of Congo, Uganda, Cameroon, Mediterranean countries like Italy, Turkey, Greece, India, Saudi Arabia. It becomes a global health burden in children as mortality rate of sickle cell anaemic children increased before reaching adulthood due to its serious complications. The distribution of sickle cell anaemic persons in the world is shown in Fig. 1.1. In India, the sickle cell gene was first detected in Nilgiri hills of Tamil Nadu in 1952 which is now widely spread over central, west and north-eastern parts of India [12]. The frequency of sickle cell anaemia is highest in tribal communities of central and southern India. The states with higher prevalence rate are Madhya Pradesh, Chhattisgarh, West Bengal, Odisha, Gujarat, Tamil Nadu, and Kerala [63]. The prevalence of the disease in India is illustrated in Fig. 1.2. In Assam, tea garden worker are mostly affected by this disease [64]. It is estimated that approximately 5,200 newborns are affected by sickle cell anaemia each year. The socio-economic ethnic people like scheduled tribes has 5% to 34% prevalence rate of sickle cell defective gene. Tribal community in India forms a very large populations of about 180 millions with people carrying sickle cell gene 18 million and people having sickle cell anaemia 1.4 million [65, 66]. So this depicts a huge health burden of India. As sickle cell anemia is an incurable inherited disorder, early detection is very much important to prevent arising complicated health issues. The mortality

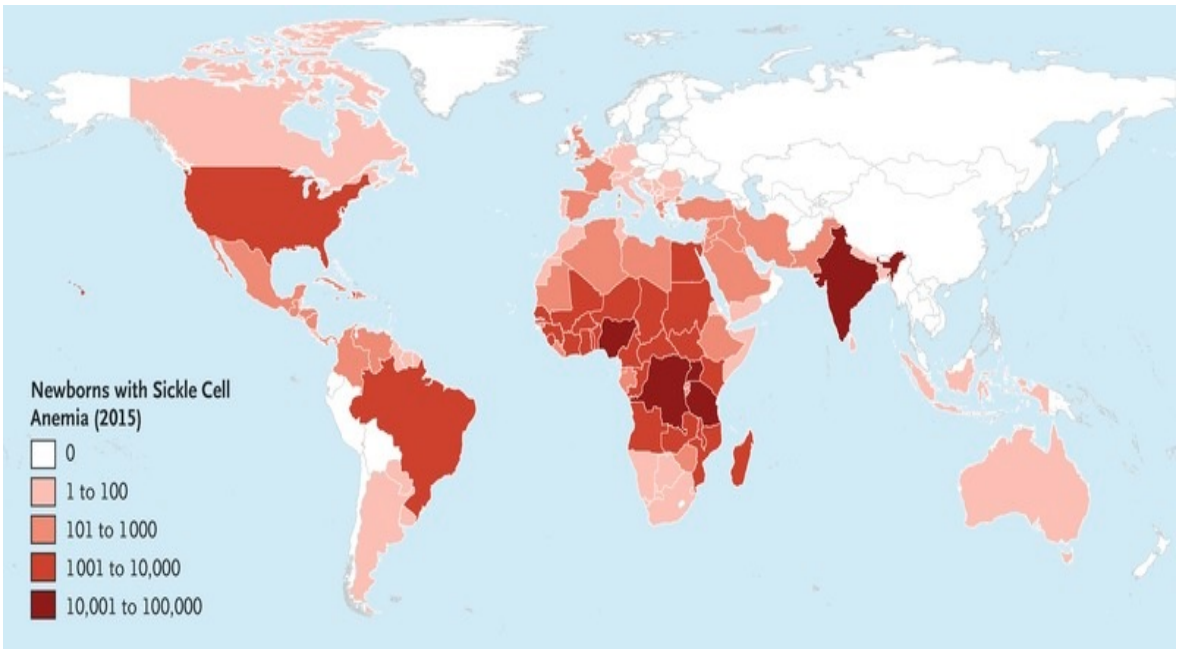


Fig. 1.1 Spread of sickle cell anaemia worldwide (Source: F. B. Piel *et al.* Sickle cell disease, The New England Journal of Medicine **376**, 1561-1573, 2017)

of children under age 5 can be decreased if proper screening of newborns is done at birth time. So diagnosis of the disease at early stage is playing a vital role for proper treatment. In many parts of world, most of the people born with SCD die before diagnosis in early childhood. There are very limited study and monitoring programe on SCD which leads to become SCD a global health burden.

There is certain need of comprehensive health program to tackle sickle cell disease and effective diagnosis technique plays a crucial part for attaining lower prevalence of the disease. An extensive newborn screening system is very much important for this under recognised disease and this screening programme is well achieved with accurate and fast diagnosis testing system. These reasons motivate to study and work on automated detection of sickle cells for properly assess the burden of the disease, leading to improving the quality of life of ethnic people all over the world. Digital image processing based identification of deformed RBCs for sickle cell disease from images of blood smear provide a fast, automatic method.

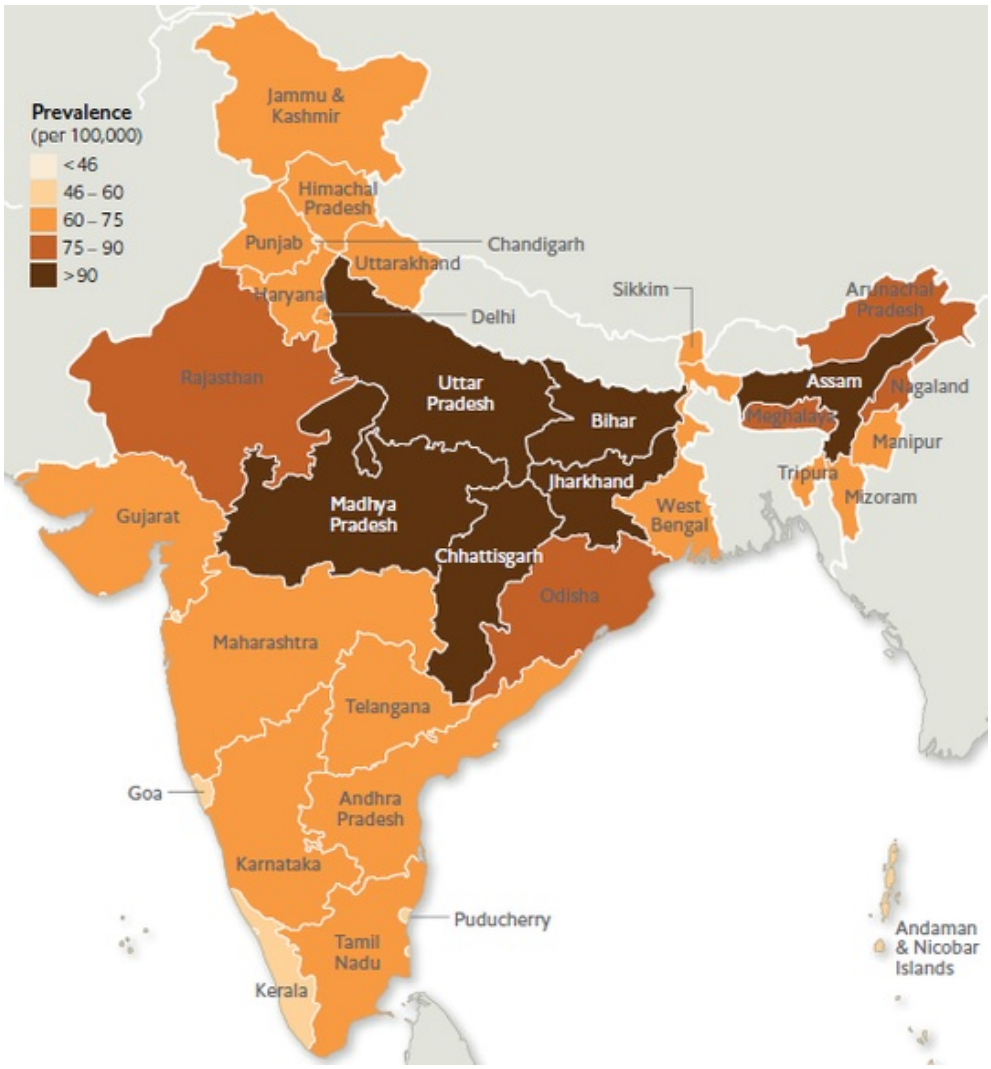


Fig. 1.2 Prevalence of Sickle cell disease in India (Source: Global Burden of Disease (GBD) "Invisible" Pain endured by millions of patients across India)

Till now, no work has been reported for sickle cell anaemia detection from blood images of anaemic patient applying adaptive thresholding based segmentation method.

## 1.4 Problem Statement

The problems for the work are formulated as follows:

1. Design a simple, automated image processing based sickle cell anaemia detection system with faster computational time and high accuracy rate.

2. Propose a hybrid segmentation procedure along with supervised machine learning techniques for classification of the disease accurately.

## 1.5 Contribution

The major contributions of the proposed work reported in this thesis include the following:

1. A local adaptive thresholding based segmentation technique is proposed for normal detection of RBCs and sickle RBCs from microscopic images. Four different adaptive thresholding methods are studied and applied for efficient detection process. A study on comparative analysis is also carried out for these four methods.

2. A hybrid segmentation based classification system of sickle cells detection is proposed. This proposed system classifies normal and abnormal RBCs by combining four adaptive thresholding along with fuzzy C-means method for segmentation. Here, two supervised machine learning techniques namely Naïve Bayes and K-nearest neighbor are used as classifier and their performance parameters are evaluated.

3. An ANN and SVM classifier based classification system of sickle cell is implemented using four different adaptive thresholding and active contour-based segmentation methods and compare the performances with that of Otsu's global thresholding method. In addition, the performance is also tested for the ANN classifier by use of various back-propagation methods.

4. The fourth contribution of work is the application of extreme learning machine along with hybrid segmentation method for sickle cell detection.

## 1.6 Organization of the thesis

The chapters of the thesis are organized as follows:

The second chapter describes about the theoretical background related with the detection of sickle cell anemia. It includes about sickle cell disease, theory of segmentation methods used in the work and different classifiers applied for sickle cells detection.

The third chapter reports the first methodology used for normal RBCs and sickle RBCs segmentation and identification. Here, normal and sickle RBCs are segmented using Niblack's adaptive thresholding method and comparative analyses between different adaptive thresholding methods are discussed with experimental results.

In the fourth chapter, second methodology for classification of sickle cells using Naïve Bayes and K-nearest neighbor classifier is described with performance parameters of both the classifiers. For segmentation of normal and abnormal cells from microscopic images of anaemic patients, here a hybrid method is proposed consisting of adaptive thresholding and fuzzy C-means technique.

The fifth chapter discusses about the third method of sickle cell detection which is based on ANN and SVM classifier. The hybrid segmentation method combining adaptive thresholding and active contour is applied. Here, Otsu's global thresholding method is also combined with active contour segmentation and results are compared.

In the sixth chapter, fourth method for detection is reported using ELM classifier and hybrid segmentation method.

In the seventh chapter, the future scopes about the reported work are discussed in details. The conclusion of the thesis is also illustrated in this chapter.

# 2

# Theoretical Background

This chapter gives a brief overview about the symptoms, causes, inheritance pattern of sickle cell disease. Different types of sickle cell disease are also discussed. After that different image segmentation techniques used for identification of normal and sickle RBCs in the work are explained. Our proposed sickle cell detection system is formulated on different supervised machine learning methods for classification purpose. The theoretical background of these supervised learning methods are also discussed in this chapter.

## 2.1 Sickle cell anaemia

Anaemia is a serious blood disorder which is mainly caused due to deficiency in the number of healthy red blood cells (RBCs) in the human body. People of all over the world with all ages are affected by anemia. There are different types of anaemia with different causes which can range from mild to severe. The main types of anaemia based on its causes are as follows:

- Anaemia due to loss of RBCs
- Anaemia due to destruction of RBCs
- Anaemia due to production of faulty or less number of RBCs

Many reasons or conditions are included for anaemia due to loss of RBCs like ulcers and gastritis in stomach, heavy menstruation periods in women, cancer in stomach etc. The second type of anaemia i.e. due to destruction of RBCs before their life span also known as haemolytic anemia includes autoimmune disorders, blood cancers, sickle cell anaemia, thalassemia and spleen enlargement. The circumstances that are associated with third type of anaemia include sickle cell anemia, iron deficiency anaemia, vitamin deficiency anaemia, thalassemia and aplastic anaemia. In aplastic anaemia, the human body does not produce enough amount of new blood cells due to the failure of bone marrow. The stem cell inside the bone marrow stops to make new white blood cells, red blood cell and platelets. Aplastic anemia may have symptoms like dizziness, fatigue, persistent infections, shortness of breath etc. Iron deficiency anaemia occurs due to insufficient amount of iron in the human body that helps to make red blood cells. This type of anaemia is more common in pregnant women, premature newborns, children, poor diets people etc. Thalassemia is inherited type of anaemia where body produce less numbers of haemoglobin affecting proper functioning of the red blood cells. The reason for thalassemia is mutation of genes those are involved in production of haemoglobin. In vitamin deficiency anaemia, human body have less numbers of normal RBCs due to deficiency of vitamins — B12, folic acid, and vitamin C. Yellowing

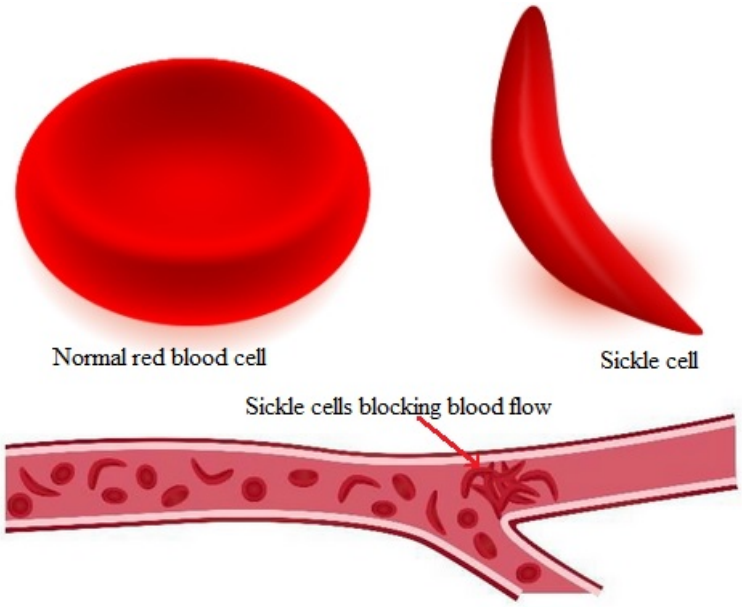


Fig. 2.1 Normal and sickle red blood cell

of skin, tiredness, rapid heartbeat, coldness of hands and feet etc. are the main symptoms of vitamin deficiency anaemia.

Sickle cell anaemia is inherited haemolytic anaemia which is caused due to defect in the hemoglobin present in the RBC. The defect in the haemoglobin causes change in shape of RBC into abnormal crescent shape shown in Fig. 2.1. This irregular shape of RBC restricts the blood flow in small blood vessels and creates serious illness. The normal RBCs are round in shape and flexible. Therefore, these cells provide oxygen to different parts of the human body by moving through small blood veins. These circular-shaped RBCs changed their shape into crescent or sickle and become rigid in case of sickle cell disease thereby preventing the normal flow of blood in the body. The defect in the gene of oxygen-carrying protein molecule haemoglobin in the blood is responsible for the change of shape of these RBCs. The defective haemoglobin is known as sickle haemoglobin or haemoglobin S (HbS). There are two $\alpha$ and two $\beta$-globin chains in the normal haemoglobin also known as haemoglobin A (HbA). In sickle cell disease, the $\beta$-globin chain of the haemoglobin becomes defective due to point mutation. Fig. 2.2 shows the mutation in $\beta$-globin chain of haemoglobin. At

the $6^{th}$ position, the hydrophilic amino acid-glutamic acid is changed to hydrophobic amino acid-valine [67].

Based on different ways of mutations in the two $\beta$-globin chains of the haemoglobin, there are different types of sickle cell disease as follows.

(a) HbAS: In this type, a person with HbAS inherits one normal gene HbA from one parent and one defective gene HbS from another parent. These persons are also called sickle cell trait (SCT). They do not show any symptoms living a normal life. However, they can pass the HbS gene to the next generation.

(b) HbSS: This type is called sickle cell anaemia, where a person gets defective gene HbS from both parents. This is most frequent and serious type of sickle cell disease.

(c) HbS $\beta$-thalassemia: In HbS $\beta$-thalassemia, one parent passes a HbS gene and the other parent passes $\beta$-thalassemia gene to their children.

(d) HbSC: This occurs when a person inherits a HbS gene from one parent and a defective hemoglobin-C gene from another parent.

(e) HbSD: In HBSD, a person inherits a HbS gene and a defective gene HbD from the parents.

(f) HbSE: Here, a person gets a defective gene HbE along with a HbS gene from the parents.

(g) HbSO: In this type, a person gets a defective gene HbE along with a HbS gene from the parents.

These different types of sickle cell diseases are pictorially depicted in Fig. 2.3 [68, 69].

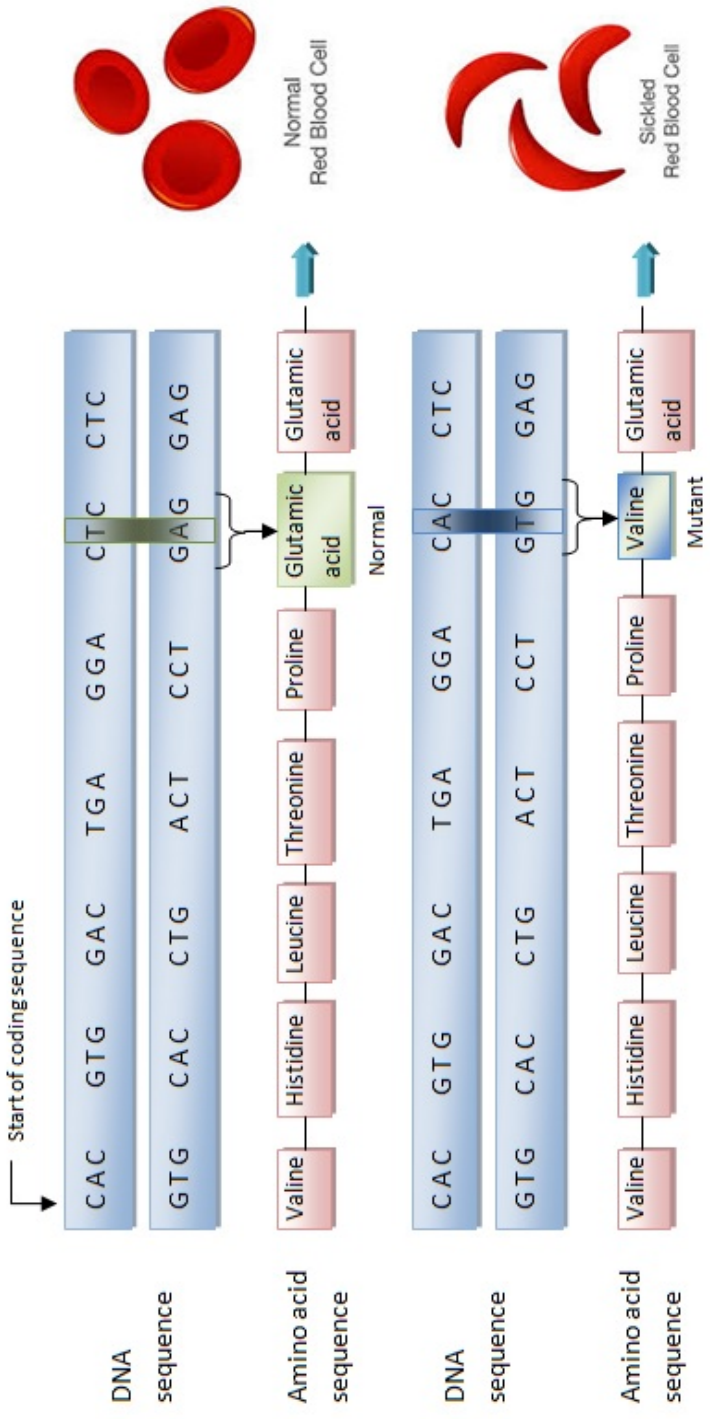


Fig. 2.2 Mutation in sickle cell disease

The inheritance pattern of sickle cell disease is shown in Fig. 2.4. In the first case, the risk for children having SCD is 100% if both parent have SCD. The possibility of having SCT is 100% if one parent has normal gene and other has SCD. If both parents are SCT, then the probability of having child with SCT is 50%, child with SCD is 25% and child with unaffected gene is 25%. The chances of having children born normal or with SCT is equal i.e. 50% if one parent is normal and other parent is SCT. If one parent is SCD and other is SCT then the probility for children having SCD or SCT is 50%.

## 2.2 Image segmentation

Image segmentation is an important step of processing an image in which an image is splitted into number of subgroups (set of pixels) or parts to reduce the complexity of an image for better analysis. Segmentation process divides an image mainly based on sudden changes in intensity or similarity of regions according to certain attributes of the image. It helps to analyze the objects present in the image by differentiating region of interest as foreground and rest part as background. It is the crucial step in all fields of image processing application without which further processing would not be possible. Segmentation has many application in various fields like object detection in satellite images, faulty red blood cells, cancer cells, tumour detection in medical imaging, fingerprint, face, iris recognition in security purposes, agriculture, manufacturing etc [70, 71]. Figure 2.5 shows the various segmentation methods in image processing for SCA detection. The main types of image segmentation techniques are as follows:

- Thresholding based segmentation
- Region based segmentation
- Edge based segmentation

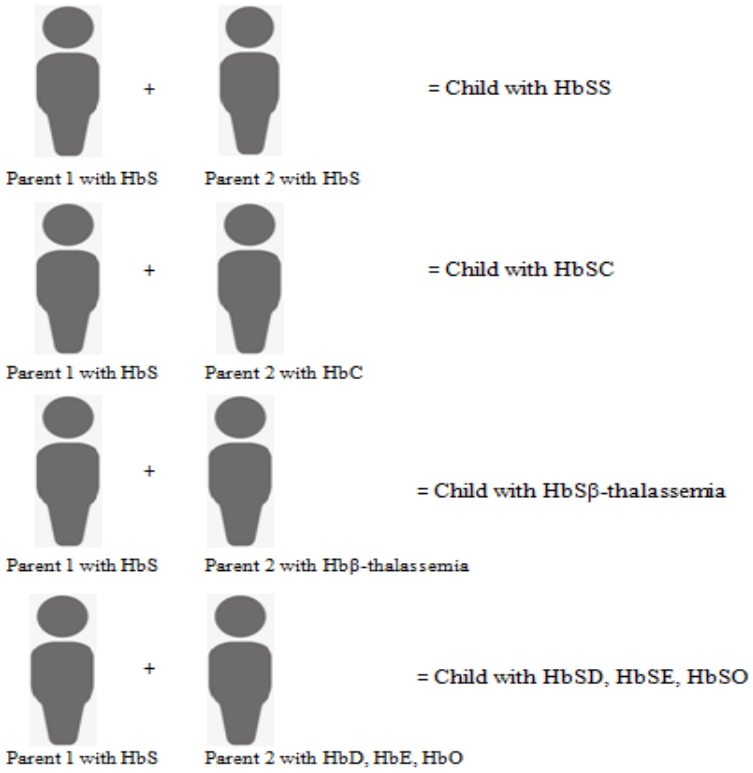


Fig. 2.3 Types of sickle cell disease

- Clustering based segmentation
- Neural network based segmentation

### 2.2.1 Thresholding based segmentation

Thresholding is the simplest image segmentation method used widely in almost every image processing fields. This method segments the image by studying the intensity levels of pixels with a specified level called threshold. Thus it separates the image into foreground with black and background with white values. There are two types of thresholding techniques: global and adaptive thresholding.

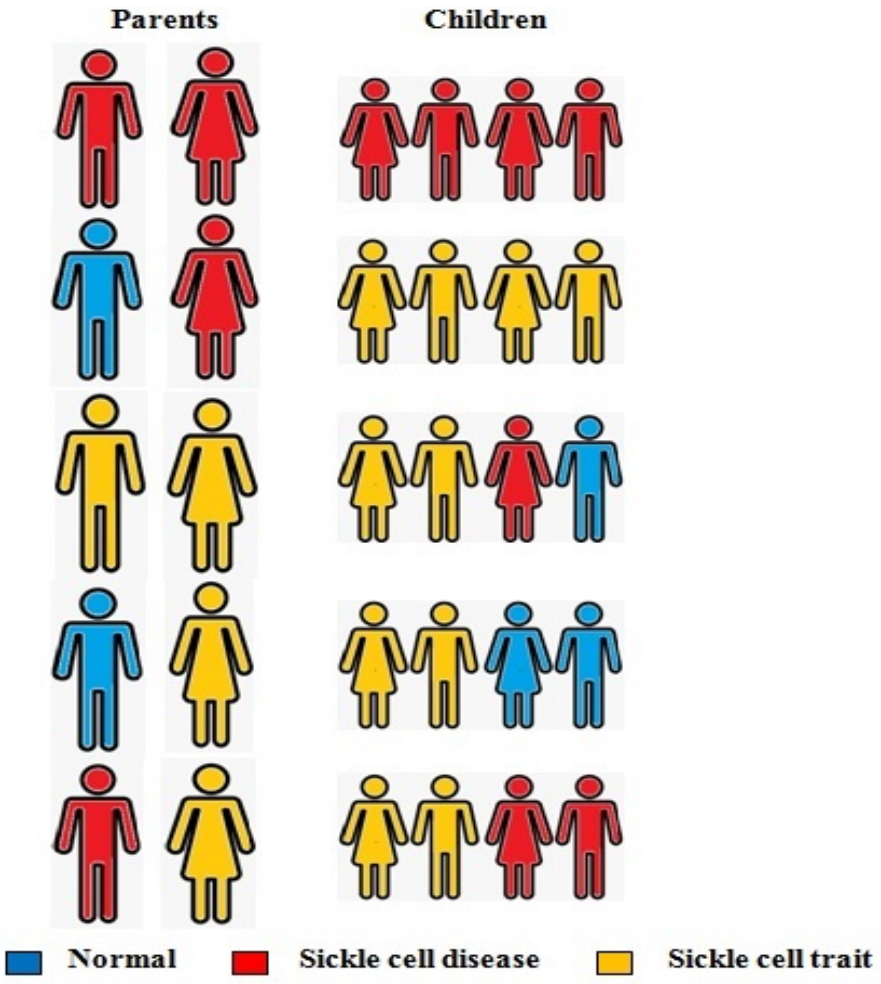


Fig. 2.4 Sickle cell inheritance

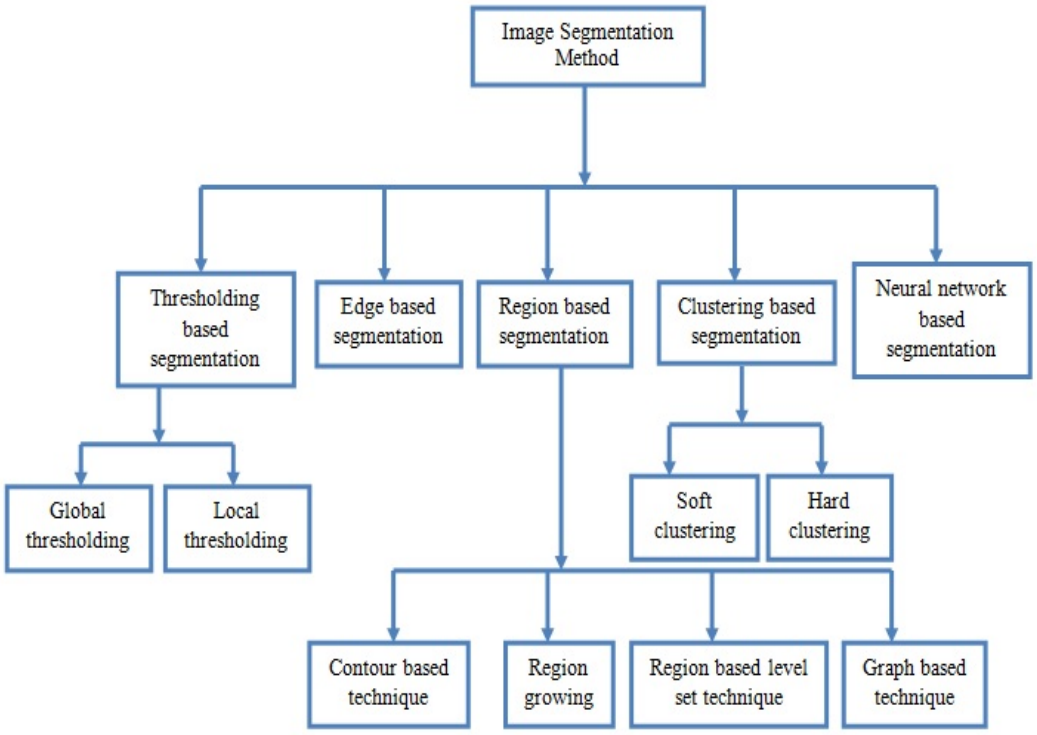


Fig. 2.5 Segmentation methods for sickle cell anaemia

**Global thresholding**

In global thresholding only one threshold value is used for the whole image for segmenting the image. This fixed threshold value generates a binary image. Global thresholding replace black (zero) to the pixels below the threshold value and white (one) to the pixels above the threshold value. This thresholding method is useful for images having bimodal histogram. Suppose $f(x,y)$ is an image having two distinct histogram with threshold value $T$ which is applied to segment the object from the background then final segmented image(binary) is given by:

$$h(x) = \begin{cases} 1, & (x,y) \geq T \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

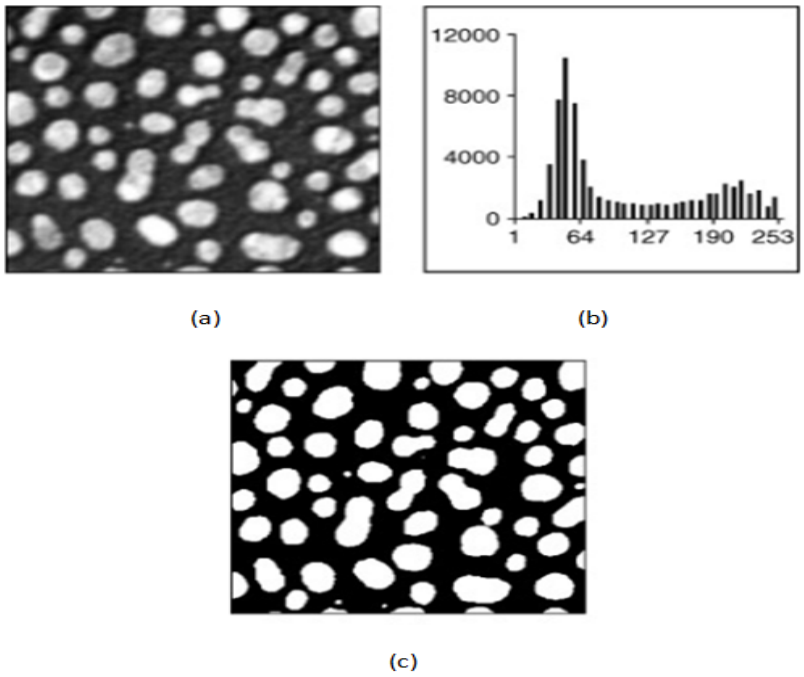


Fig. 2.6 (a) Original image (b) Histogram of image (c) Result of global thresholding

An example of global thresholding is shown in Fig. 2.6 where the image is segmented using a single threshold value $T = 127$. The pixels of the image having intensity value lower than 127 are converted to black and pixels having intensity value higher than 127 are converted to white in the thresholded image.

**Otsu's thresholding:**

There are some limitations of manually setting a predetermined threshold value. Therefore automatic selection of threshold value is important for image with changing lighting conditions. Otsu's thresholding proposed by Nobuyuki Otsu implements single threshold value automatically by minimizing the intra-class variance in intensity or maximizing inter-class variance.

#### Adaptive thresholding

The global thresholding method performs inferior for images with varying conditions of light, contrast or colours in the image. The results for some areas are not satisfactory for such types of images. Adaptive thresholding is used for these types of images to overcome the disadvantages of global thresholding. The whole image is splitted into numbers of smaller sub-images in adaptive thresholding and for each of these sub-images different threshold values are evaluated. The calculation of the threshold value for sub-images is based on statistical tools like mean and/or standards deviation etc. After calculating threshold values of sub-images, the final segmented image is obtained by merging the sub-images. Figure 2.7 shows an example of adaptive thresholding.

There are many popular adaptive thresholding methods proposed by researchers used in many image processing fields for segmentation. These are namely Niblack's method, White and Rohrer's method, Bernsen's method, Yanowitz and Bruckstein's method, Yasuda, Dubois and Huang's method, Sauvola's method, NICK's method [72].

### 2.2.2 Clustering based segmentation

Clustering is an approach where data points are grouped together based on similar characteristics or features like distance, intensity, connectivity. Data points which are in same cluster should have same properties, while data points in different clusters should have different properties. Therefore the goal of clustering is to divide the data into groups having same properties. Figure 2.8 shows an example of clustering. There are two types of clustering. They are: hard clustering and soft clustering. In hard clustering, each data points should belongs to only one cluster. In soft clustering, each data points can belongs to more than one cluster. Here each data point is assigned with a probability or likelihood to be in a cluster [73]. Also there are different types of clustering algorithms like partition clustering, distribution-based clustering, hierarchical clustering, fuzzy clustering, density-based clustering etc.

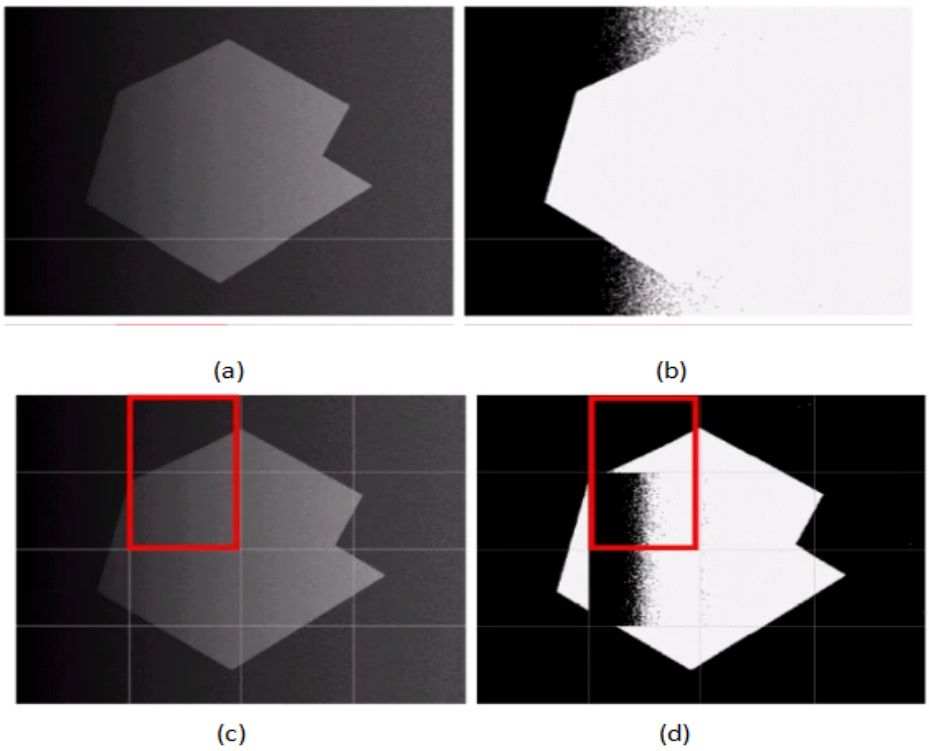


Fig. 2.7 (a) Original image (b) Result of global thresholding (c) Image subdivided into individual sub-images (d) Result of adaptive thresholding

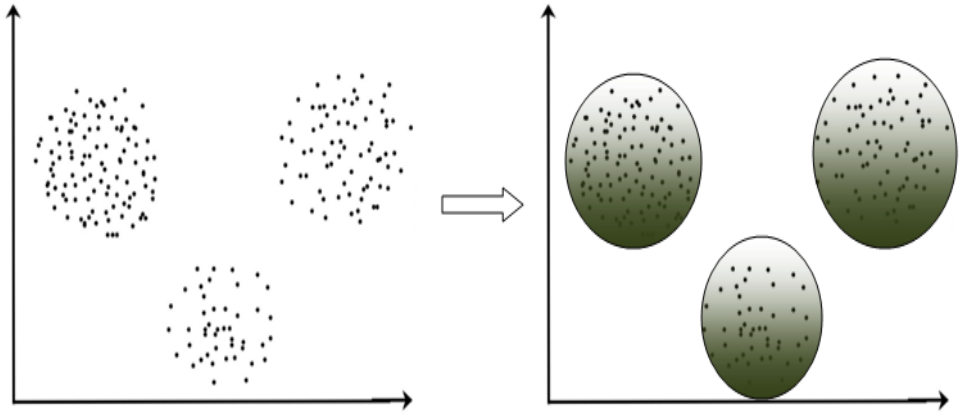

Fig. 2.8 Example of clustering

### Fuzzy C-means clustering

It is a popular soft clustering technique with data points can be associated with two or more clusters simultaneously. In fuzzy C-means, membership values are assigned to each data point for each of the cluster center. These membership values specify the extent to which data points are associated with each cluster. The lower membership value of data points on the edge of a cluster are lesser degree to be in the cluster than the data points which are in the cluster center [74]. The algorithm has the following steps:
Let $Y = y_1, y_2, y_3..., y_m$ be the set of data points and $V = v_1, v_2, v_3..., v_n$ be the set of centers.

1. First select $n$ cluster centers.

2. Evaluate fuzzy membership $\mu_{pq}$ using formula:

$$\mu_{pq} = \frac{1}{\sum_{l=1}^{n} (\frac{d_{pq}}{d_{pl}})^{\frac{2}{c-1}}} \tag{2.2}$$

   where, $d_{pq}$= Euclidean distance between $p^{th}$ data and $q^{th}$ cluster center.

3. Compute fuzzy centers $v_q$ using formula:

$$v_q = \frac{\sum_{p=1}^{m} (\mu_{pq})^c y_p}{\sum_{p=1}^{m} (\mu_{pq})^c} \tag{2.3}$$

   where, c= Fuzziness parameter.

4. The goal of fuzzy C-means clustering is to minimize the objective function given by:

$$F = \sum_{p=1}^{m} \sum_{q=1}^{n} (\mu_{pq})^c \|y_p - v_q\|^2 \tag{2.4}$$

Repeat second and third steps until the value of $F$ become minimum or $\|U^{k+1} - U^k < \beta\|$. where, $U = (\mu_{pq})_{m*n}$= Fuzzy membership matrix, $k$=Iteration step, $\beta$= termination criterion between 0 and 1 [75].

### 2.2.3 Region based segmentation

In region based segmentation, a region is formed as a set of pixels having same properties which can be in respect of gray levels of pixels, colour of pixels, textures etc. There are two main types of region based segmentation. They are: a) Region growing technique and b) Region splitting and merging technique. In region growing technique, neighbouring pixels are added to the seed points based on similarity forming a large region. The seed point is selected arbitrarily in the initial stage. The process of adding pixels to seed points is stopped when all pixels of the image belongs to some region. In split and merge technique, whole image is first subdivided into homogeneous regions and after that merge neighbouring regions with similar properties.

#### Active contour-based segmentation

When first employed by Kass *et al.* in 1988 [76], active contour model is used for image segmentation and boundary tracing. This is a region based iterative boundary detection technique starting with an initial contour for target boundary. The contour is then changed so that it reaches the final desired boundary of the object with some defined criteria. Active contour model uses the energy constraints and forces for partitioning the region of interest in the image. It creates a contour or parametric curve for borders of the object. Various contour methods using internal and external forces are employed for determining the curvature of the model. The energy function is linked with the curve of the image. The sum of forces applied to the image for controlling the contour position on the image described the external energy. Internal energy is used to control the deformable changes in the image. The desired contour

will be defined by the least energy function. This model is proved to be very effective for segmentation of variety of medical images [77, 78].

## 2.3 Supervised machine learning

Supervised learning is a class of machine learning in which computers or machines are trained with labelled dataset and machines produce output based on those training data. The labelled dataset consist of some input data are marked with correct output. These labelled dataset are applied to machine for learning and to predict the outcomes correctly. So the machine can predict the outputs on the basis of prior knowledge or experience. The algorithms of supervised learning are categorized as classification and regression model. There are variety of applications of supervised learning algorithms on real field world including speech recognition, sentiment analysis, forecasting of weather, stock market prediction, handwritten text detection, spam detection, object recognition etc.

### 2.3.1 Naïve Bayes classifier

This classifier is one significant type of supervised learning technique based on Bayes theorem. Known after the name of Thomas Bayes, it works on the assumption that the features are not statistically dependent. Being a probabilistic method, it assumes the existence of a specific feature in a class as independent of having any other features. One of the important advantages of this method is that, for training purpose it requires a small dataset. Parameters required for classification can be estimated in this method [79, 80]. This method aims at finding out the posterior probability.

Applying Bayes rule, we can calculate the probability of posterior of $n$ size dependent feature vector $X$.

$$P(c_j|X) = \frac{P(X|c_j) \times P(c_j)}{P(X)} \tag{2.5}$$

where $X = (x_1, x_2, x_3, \cdots\cdots, x_n)$

$P(c_j|X)$= The class $j$'s prosterior probability with a predictor $x$

$P(X|c_j)$= Probability of $x$ given a class $j$.

$P(c_j)$= Prior probability of class $j$.

$P(X)$= Prior probability of predictor or Evidence.

Since in Naïve Bayes algorithm, conditional probabilities of the attributes $x$ are assumed to be statistically independent, decomposing the probability of likelihood to product terms, it is estimated as

$$P(X|c_j) = P(x_1|c_j) \times P(x_2|c_j) \times P(x_3|c_j) \times \cdots P(x_n|c_j) \tag{2.6}$$

Therefore probability of posterior of Eq. (2.5) becomes

$$P(c_j|X) = \frac{P(x_1|c_j) \times P(x_2|c_j) \times P(x_3|c_j) \times \cdots P(x_n|c_j) \times P(c_j)}{P(x_1) \times P(x_2) \times P(x_3) \cdots \times P(x_n)} \tag{2.7}$$

The denominator remains constant for a given input, so Eq. (2.7) can be written by removing denominator.

$$P(c_j|X) \propto \prod_{i=1}^{n} P(x_i|c_j) \times P(c_j) \tag{2.8}$$

For classifier algorithm, probability of known set of inputs for all possible values of class variable is calculated and output with maximum probability is picked up [75]. This can be expressed mathematically as:

$$c = \arg\max_{c_j} \prod_{i=1}^{n} P(x_i|c_j) \times P(c_j) \tag{2.9}$$

### 2.3.2 K-nearest neighbor classifier

The k-nearest neighbors (KNN) algorithm is one amongst the most popular supervised machine learning method which has been extensively experimented for classification, data mining, regression and pattern recognition purposes. This non-parametric algorithm utilises 'feature similarity' for predicting the new data points values. It means, a new class will be assigned to the new data points based on their closeness of similarities with the points of the training set. The algorithm has the following steps.

1. Loading of the training as well as well as test dataset.
2. Select the values of the nearest data points $k$ ($k$ is a positive integer, typically small)
3. For all the points in the test data, calculate their distance from each row of training data.
4. Now, sort them in ascending order based on the distance value.
5. Top $k$ rows are to be chosen from the sorted array.
6. Based on the most common class of these rows, the test point will be assigned a class [75].

### 2.3.3 Artificial neural network

One of the finest tools in machine learning is ANN, which can replicate the things the way human learn things. In other words we can say that ANNs are human brain inspired systems. An ANN consists of hundreds of processing elements, called neurons. These neurons have large weighted connections between them, forming neural structure and are arranged in layers. Each connecting neurons in ANN determine the computational power of the network. Each neuron is made up of input, output and the transfer function. The transfer function

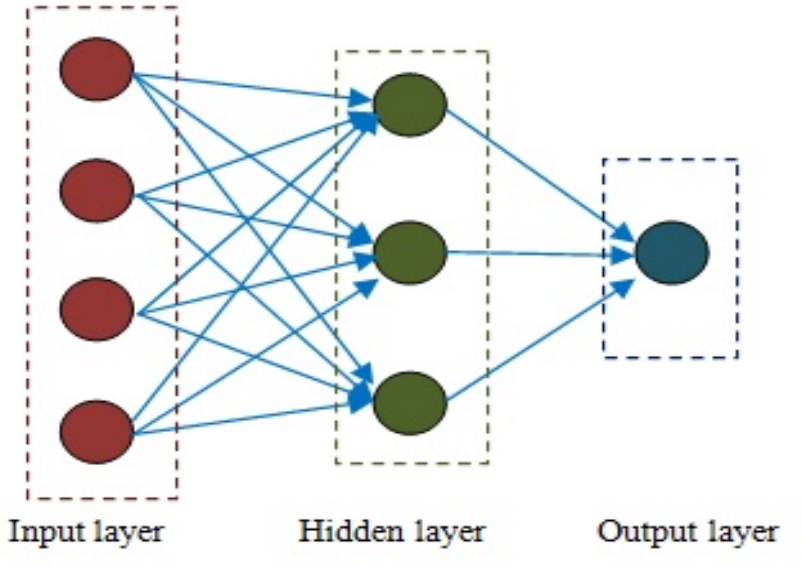


Fig. 2.9 Different layers of ANN

determines the behaviour of the network. Neurons get inputs from various sources and the weighted sum of these inputs is the activation of the neurons. The stimulating signals for activation is moved through the transfer function and generate an output of the neuron [81, 82].

The most general type of ANN is the feed forward neural network and its architecture is shown in Fig. 2.9. They have multiple neurons arranged in 3 interconnected layers: input, intermediate and the output layer. The input layer nodes just transfer data/information to the next level. The computation is done in the intermediate or the hidden layer which may be single or multiple. After computations they pass the weights to the following layer. Output layer employ an activation function that maps the data to desired output format. The main goal of ANN is to make a platform that can execute various computational tasks faster than the traditional systems. The functions ANN can handle include classification, pattern recognition, function approximation, optimization, control and data mining [83]. Other classes of NN are: Recurrent NN, Convolutional NN, Radial basis function NN, Kohonen Self Organizing NN etc.

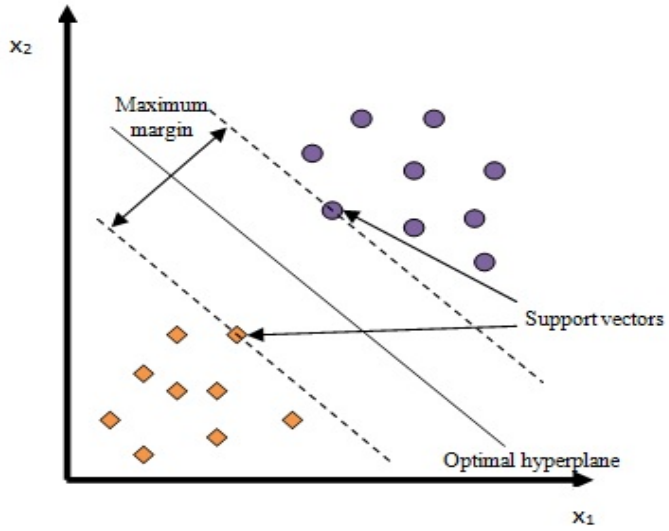


Fig. 2.10 SVM

### 2.3.4 Support vector machine

Another machine learning classifier which is frequently used in medical field is SVM. It is a fast and reliable supervised algorithm for classification and its performance is very well for a limited amount of analysing data. In SVM, all data points are plotted in p-dimensional space where $p$ is number of features considered. SVM classifier perform classification by drawing the best decision boundary in that space that separates the data points into distinct classes. This best decision boundary or line is called hyper-plane. The data points of different classes will appear on different sides of the hyperplane. The closest points of the hyperplane from both the classes are defined as support vectors and the distance between these support vectors and hyperplane is called the margin. SVM tries to maximize this margin by finding the optimal hyperplane which means the boundary separates the classes is as wide as possible as shown in Fig. 2.10 [84, 85].

### 2.3.5 Extreme learning machine

ELM is basically feed forward networks with single-hidden-layer. This networks work with arbitrarily assigned input layer weights and biases and generalised inverse of the hidden layer

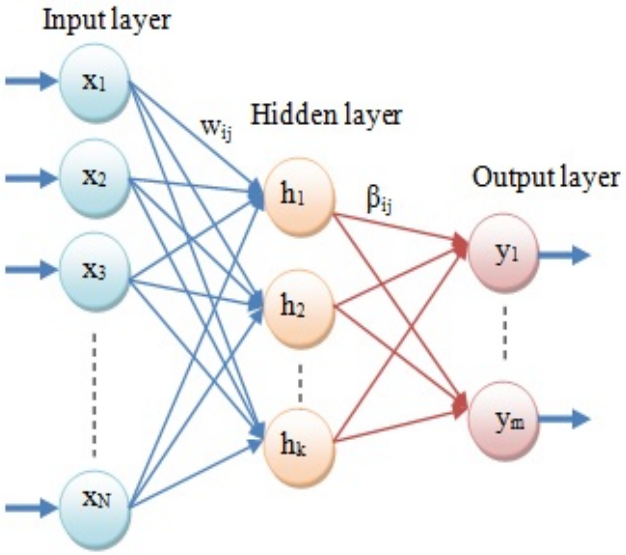


Fig. 2.11 ELM structure.

output matrix will determine the output layer weights. Figure 2.11 illustrates the layout of an ELM structure.

The output of ELM with $K$ hidden nodes is:

$$f_K(x) = \sum_{i=1}^{K} \beta_i g(w_i * x_j + b_i) \qquad j = 1, 2, ....., N \tag{2.10}$$

where, $\beta_i$ = weight of output layer of the $i^{th}$ neuron of hidden layer, $g$ = activation function, $w_i$ = weights of input layer related to the $i^{th}$ neuron of hidden layer, $x_j$ = $j^{th}$ input sample, $b_i$ = bias, $N$ = no. of training samples.

Equation 2.10 can be shorten and written as:

$$T = G\beta \tag{2.11}$$

where,

$$G = \begin{bmatrix} g(w_1 * x_1 + b_1) & \cdots & g(w_K * x_1 + b_1) \\ \vdots & \cdots & \vdots \\ g(w_1 * x_N + b_1) & \cdots & g(w_K * x_N + b_K) \end{bmatrix}_{N \times K} \tag{2.12}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_K^T \end{bmatrix}_{K \times m} \tag{2.13}$$

and

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{2.14}$$

where, $m$ = no. of output, $G$ = hidden layer output matrix, $T$ = training data target matrix.

ELM is a linear system if the parameters of hidden layer $w_i$ and $b_i$ can be assigned randomly. Then cost function is created:

$$\|G\hat{\beta} - T\| = \min_{\beta} \|G\beta - T\| \tag{2.15}$$

The values of $\beta$ is calculated by determining a least square solution as:

$$\hat{\beta} = G^{\dagger}T \tag{2.16}$$

where $G^{\dagger}$ = generalized inverse of $G$ (Moore-Penrose). Therefore, the output weights is calculated by a mathematical transformation [86, 87].

# 3

# Detection of Red Blood Cell and Sickle Cell from Microscopic Blood Images using Local Adaptive Thresholding Techniques

## 3.1 Introduction

Different segmentation methods have been applied by researchers in the medical field for the diagnosis of normal and abnormal RBCs as discussed in the literature section of Chapter

1. Thresholding based segmentation methods are used extensively in many works by the researchers [16, 17, 25, 26, 43, 50]. They mainly used global thresholding methods for SCA detection. In this chapter, local adaptive thresholding techniques are discussed and results were compared for red blood cells (RBCs) and sickle cells detection. First, Niblack's thresholding method was applied for segmentation of normal and sickle cells. The geometric parameter i.e. metric value or form factors of the RBCs are used for detection process. In the second methodology, another three adaptive thresholding methods namely Bernsen, Sauvola and NICK along with Niblack were implemented one by one on the image database and a comparative analysis was carried out.

## 3.2 Detection of sickle cells using Niblack's adaptive thresholding method

### 3.2.1 System Block Diagram

The block diagram for detection of normal and sickle RBCs from microscopic blood images of anaemic person is shown in Fig. 3.1. The block diagram consists of four steps: a) input image acquisition, b) image pre-processing, c) segmentation using Niblack's thresholding and d) detection of abnormal RBCs using form factor. Red blood cells detection is the process to differentiate between normal RBC and sickle cell in microscopic blood image [88].

#### Data acquisition and image pre-processing

In this work we have used samples collected from online medical library [89]. Figure 3.2 and Fig. 3.3 show some of sample microscopic blood images of sickle cell anaemic patients containing normal RBCs and sickle RBCs. A total of 10 such images are collected to form the

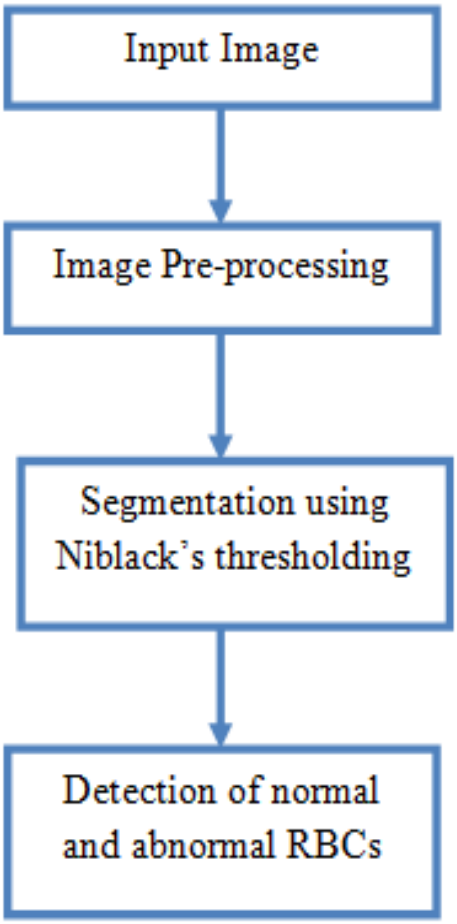


Fig. 3.1 Block diagram for detection of sickle cells

image database. These images are enhanced for further analysis. Therefore pre-processing is done before feeding into the segmentation process.

Pre-processing of image is an essential step in image processing so that the image is set for further actual processing. Image enhancement is done after removing noises present in the image. There are some artefacts and illumination issue which have been added during image acquisition; these must be removed in the pre-processing step. Filtering, debluring, histogram equalization, contrast enhancement etc. are several practises that exist for pre-processing of images. After reading the RGB input image in the MATLAB environment, it is transformed to gray scale. Thereafter the image can be filtered using a median filter to

remove the impulsive noise present in the image if necessary. Median filter is used here as this filter is useful for preservtion of edges while removing noise.

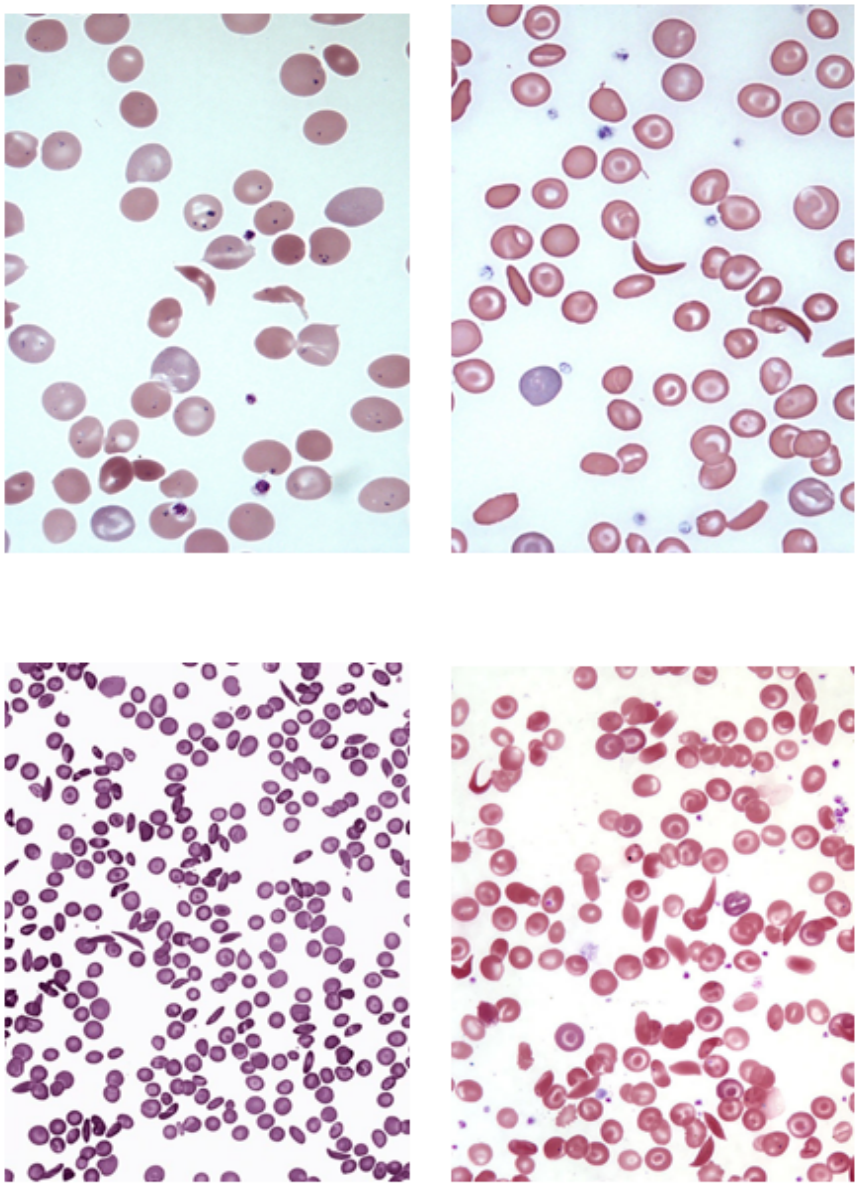

Fig. 3.2 Sample microscopic blood images of sickle cell anaemic patients

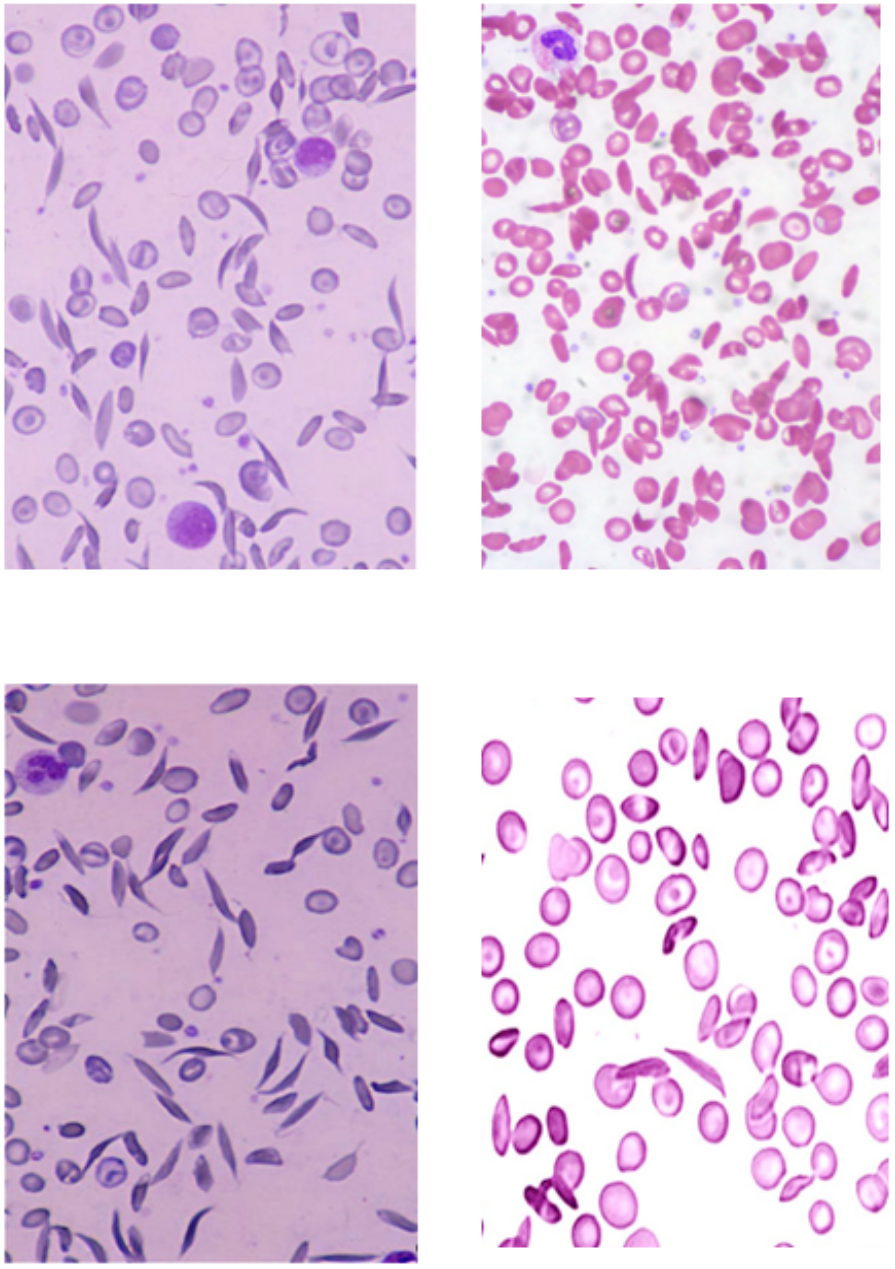

Fig. 3.3 Sample microscopic blood images of sickle cell anaemic patients

### Segmentation using Niblack's thresholding

In segmentation process the image is partitioned into different objects. In sickle cell anaemia detection, the image is separated into foreground region with the aim to isolate the normal and abnormal RBCs and background region with the plasma. Image thresholding procedure is one of the vital ways for segmentation of image that separate the image into foreground and background regions. Here, Niblack's thresholding method is applied to get the binary image from the gray level one. This effectively separates the cell image from its background.

### Niblack's thresholding

This local thresholding algorithm is based on calculation of threshold value using mean and the standard deviation by moving window about the location of each pixel. The native threshold at any given pixel $(x,y)$ can be calculated by using the equation as given in Eq. 3.1.

$$N_T(x,y) = m(x,y) + k * \delta(x,y) \tag{3.1}$$

Here, $m(x,y)$ and $\delta(x,y)$ are the local mean and local standard deviation respectively. The quality of thresholded image determines with the value of $k$ and the size of the sliding window. Based on the type of application, the size of the window varies. Segmentation result of binary image depends on the size of window [91, 92]. After Niblack's thresholding, morphological operation has been applied on the image in order to remove the distortions and to make the object smooth. The holes in the binary image are also filled which refines the border of the image. After that small objects have been removed from the binary image. The final segmented image for a sample image is shown in Fig. 3.4.

**Detection of normal RBCs and sickle cells**

Geometrical Features are used to classify normal RBC and sickle cell from the segmented images. Here we estimate each object's area and perimeter. Using these results a simple metric (form factor) is calculated that indicates the roundness of an object as given in Eq. 3.2. The value of this metric is equal to one for a circle and it is less than one for any other shape.

$$Circularity \quad or \quad Metric \quad Value = \frac{4 \times \ pi \times area}{perimeter^2} \tag{3.2}$$

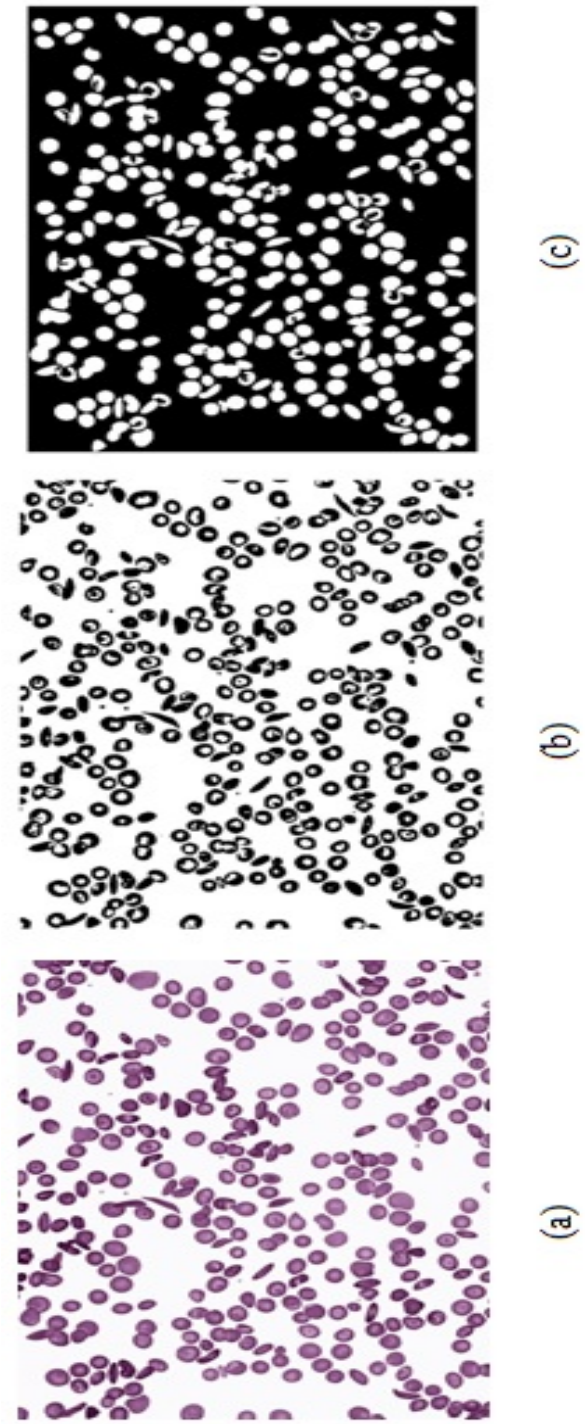


Fig. 3.4 (a) Sample image (b) Segmented image using Niblack's thresholding (c) Final segmented image

### 3.2.2 Results and Discussion

In this section, we will discuss how the proposed method is being implemented for achieving significant results in terms of detection of normal RBCs and sickle cells. The Niblack's thresholding is applied to the image data set. Then normal and abnormal RBCs are classified calculating geometrical parameter i.e. form factor of each cell in the blood images. Fig. 3.5 shows the three microscopic blood images of sickle cell anaemic persons and their corresponding segmented results after applying Niblack's thresholding and calculation of form factor to identify normal and abnormal RBCs. In Fig. 3.5 form factor determines the classification of normal and abnormal RBCs. The RBCs where form factor is nearer to 1 are detected as normal RBCs. For sickle cells value of form factor is nearer to 0.5. Thus using form factor the normal RBCs and sickle cells have been differentiated from the microscopic images.

This work describes a local thresholding algorithm that removes background of an image of blood by using local mean and standard deviation. Niblack's thresholding technique is implemented here for segmentation process on blood images. Although this method has been experimented in different applications, for detection of sickle cell anaemia it has been implemented afresh. This thresholding technique along with geometrical parameter (form factor) proves to be very much effective segmentation method in detection of normal RBC and sickle cell from microscopic blood images.

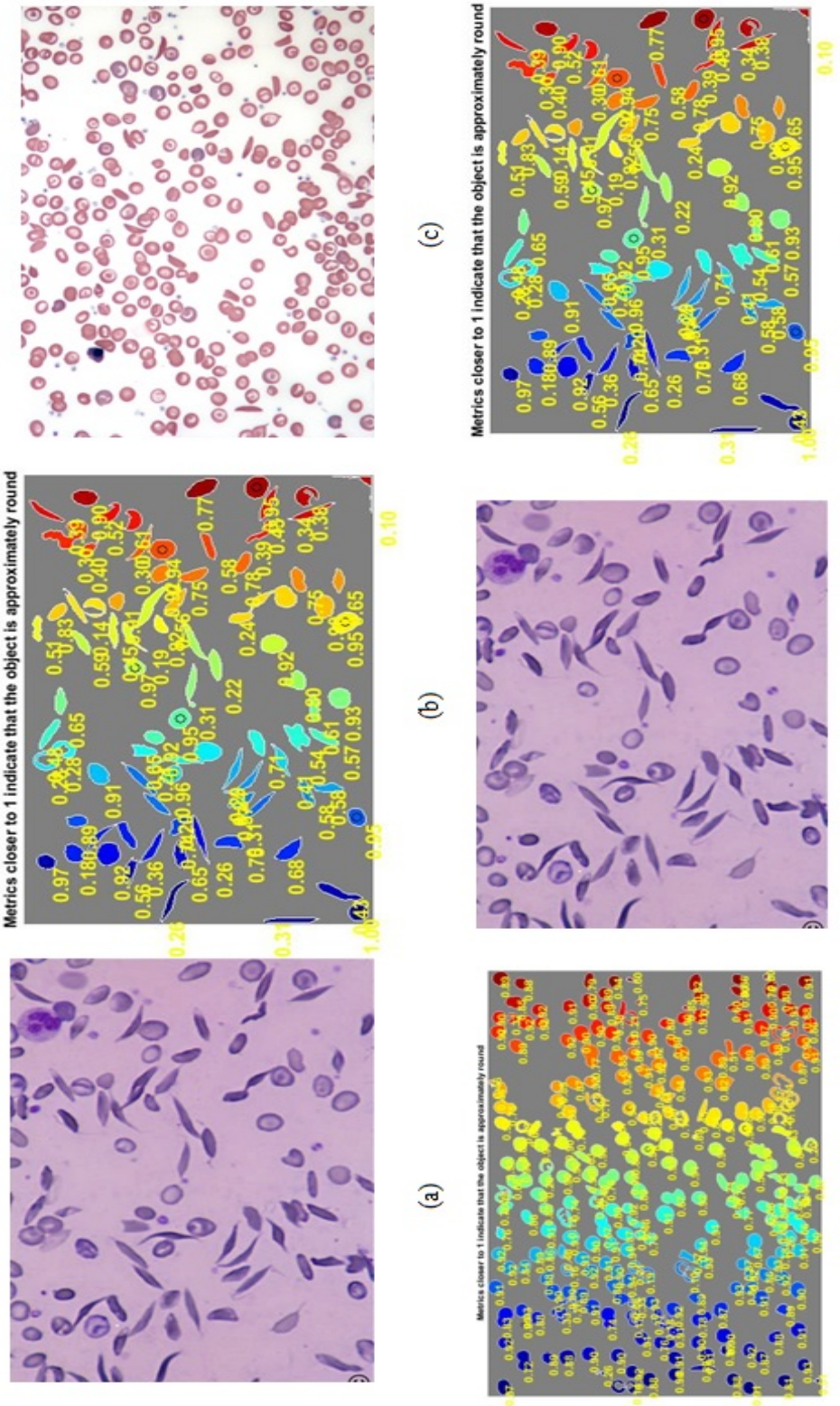


Fig. 3.5 (a) Image 1 (b) Detected normal RBCs and sickle cells of image 1 (c) Image 2 (d) Detected normal RBCs and sickle cells of image 2 (e) Image 3 (f) Detected normal RBCs and sickle cells of image 3

## 3.3 Comparative study on different adaptive thresholding techniques for detection of sickle cells

### 3.3.1 System Block Diagram

Figure 3.6 shows the block diagram of detection of normal RBCs and sickled RBCs from microscopic image of blood of person having SCD. This block diagram has four steps: a) input image acquisition, b) image pre-processing, c) segmentation using four different thresholding methods and d) detection of normal and sickle RBCs. Here the segmentation block discussed in the first methodology in section 3.2 is replaced by a segmentation block with four different adaptive thresholding methods [90].

**Image segmentation using different adaptive thresholding techniques**

In this work different adaptive thresholding methods namely Bernsen, Niblack, Sauvola and NICK techniques are applied to the pre-processed image one by one. After thresholding, different morphological operations have been applied for removal of distortions and to make the object smooth.

**(i)** Bernsen's Technique: This thresholding method is proposed by Bernsen [93] which is based on local contrast value to calculate local threshold value given by Eq. (3.3)

$$T_B(p,q) = \frac{I_{low} + I_{high}}{2} \tag{3.3}$$

Here $I_{low}$ and $I_{high}$ are the lowest and highest gray level values in a neighborhood $r \times r$ centered at $(p,q)$. The threshold value $T_B(p,q)$ depends on the local contrast value that is calculated by Eq. (3.4)

$$C(p,q) = I_{high} - I_{low} \tag{3.4}$$

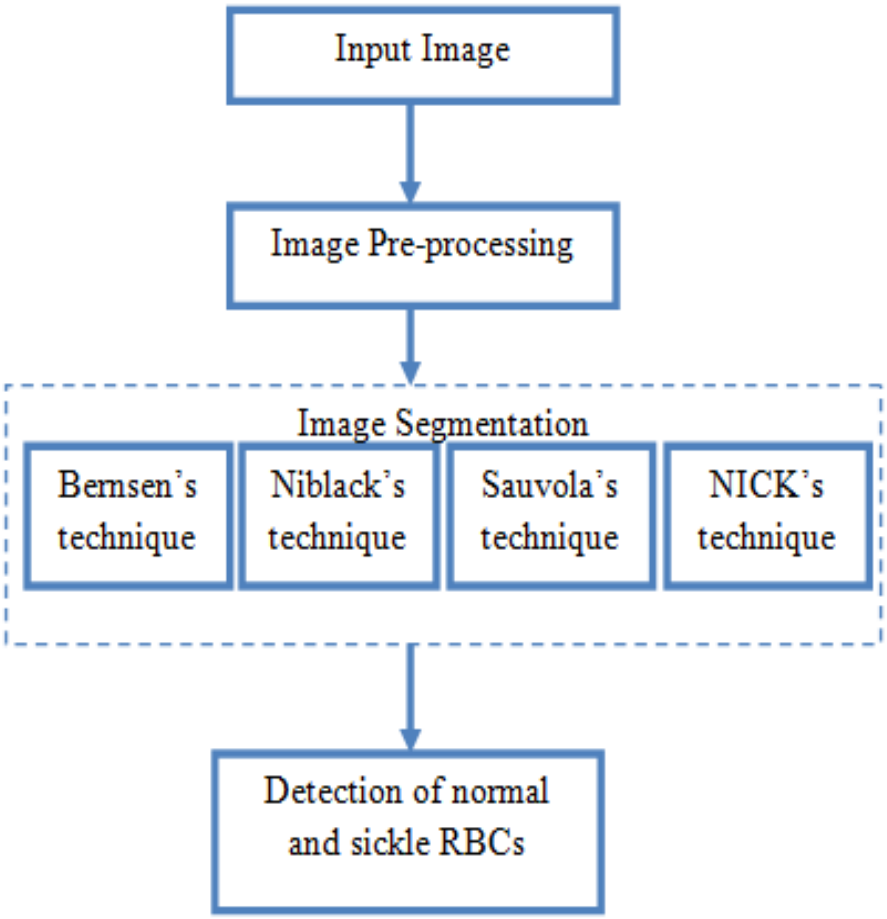


Fig. 3.6 System diagram for detection of sickle cells

If $C(p,q) \geq$ contrast threshold, then $T_B(p,q)$ is calculated by Eq. (3.3). Otherwise, the neighborhood consists only of one class, foreground or background.

**(ii)** Sauvola's Technique: This approach uses local mean, $m(p,q)$ and local standard deviation, $\delta(p,q)$ for each pixel separately to calculate $T_S(p,q)$ given by Eq. (3.5)

$$T_S(p,q) = m(p,q)[1 + P(\frac{\delta(p,q)}{S_{max}} - 1)] \tag{3.5}$$

Where, $P$ = Constant

$S_{max}$ = Max. value of standard deviation

The threshold value $T_S(p,q)$ becomes approximately equal to $m(p,q)$ for a high contrast area in an image since $\delta(p,q) \sim R$ whereas for low contrast area $T_S(p,q) <$ mean value and remove the dark portions in the background [94].

**(iii)** Niblack's Technique: The local threshold value $T_N(p,q)$ is calculated by the Eq. 3.6 which was proposed by W. Niblack [91].

$$T_N(p,q) = m(p,q) + P * \delta(p,q) \tag{3.6}$$

Here, $P$ = Constant in the range [0 to 1].

The $P$ assessment and the size of the window determine the thresholding quality in the segmentated image.

**(iv)** NICK's Technique: In this technique, the threshold value is determined from the Niblack's method as given in Eq. 3.7

$$T_{NICK}(p,q) = z + P\sqrt{\frac{(\sum b_i^2 - z^2)}{MN}} \tag{3.7}$$

Here, $P$ = Constant in the range [-0.1 to -0.2]

$z$ = Avg. level of intensity

$b_i$ = Intensity of pixel

$MN$ = Total counts of pixels present in the image [95].

### 3.3.2 Results and Discussion

The image dataset consisting of 10 samples are fed to the above mentioned thresholding techniques and detection results are studied for comparative analysis. Figure 3.7 shows one of the input sample image taken from the database used for the work and its segmented version after applying the techniques of Bernsen, Niblack, Sauvola and NICK. These figures specify the value of the metric of each cells present in the blood image to classify them into normal and abnormal RBCs.

Table 3.1 shows the overall accuracy of different thresholding techniques which are calculated after applying them to the whole image dataset. From the table, we can say that, Niblack's technique gives comparatively better result for detection of sickle cells. Table 3.2 shows the computational time of all the techniques. From computational time point of view, NICK's technique proves to be the best.

Here we have discussed about four different adaptive thresholding methods for detection purpose of normal and abnormal cells in sickle cell anaemia. Although these adaptive techniques have been used for segmentation purpose, this work focuses mainly on applying these tools for detection of sickled RBCs. From the result, we can say that out of all four thresholding methods, Niblack's technique gives highest accuracy of 96% for sickle cell detection. Again from computational time point of view, NICK's technique performs in the lowest time of 1.73 seconds.

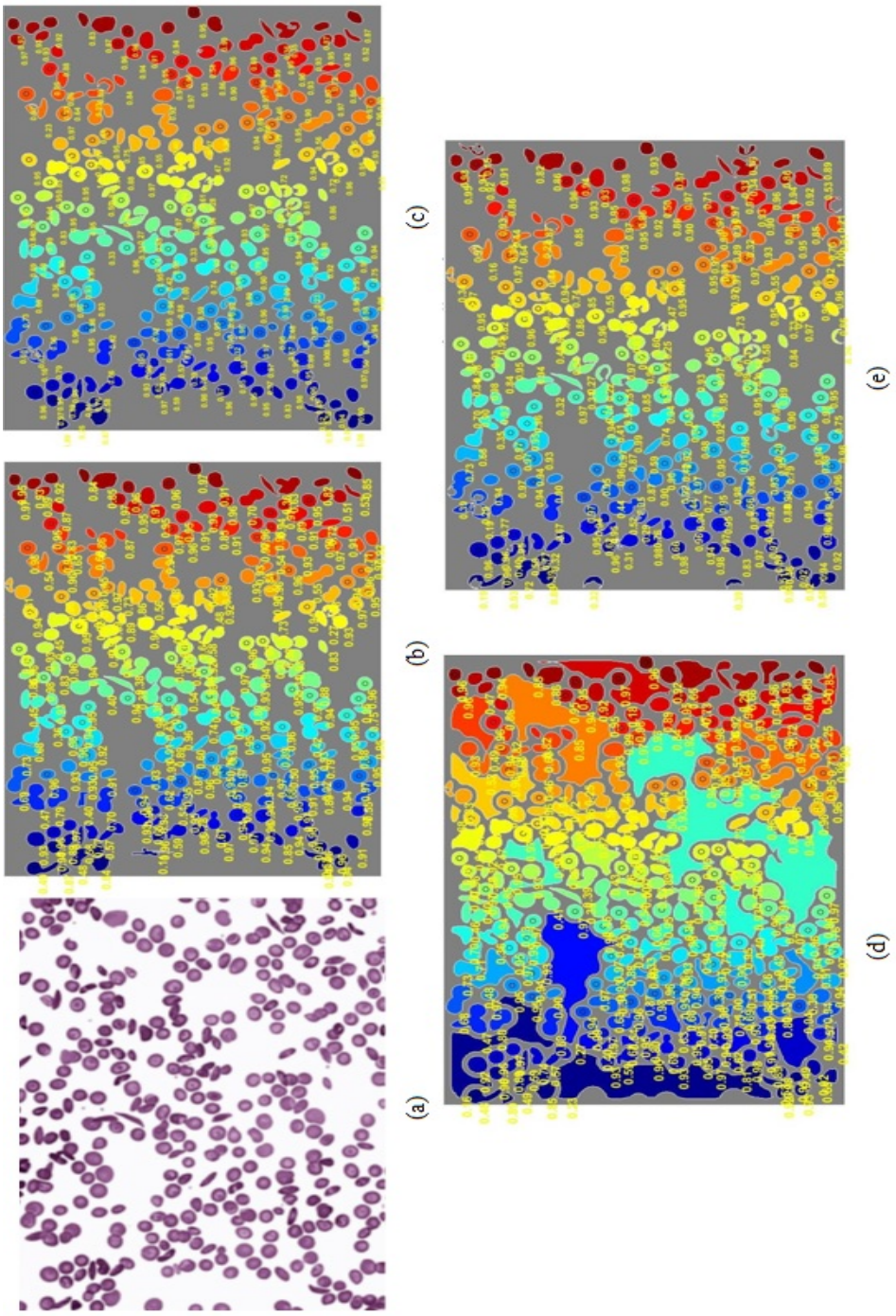


Fig. 3.7 (a) Sample image (b) Segmented image using Bernsen's thresholding (c) Segmented image using Niblack's thresholding (d) Segmented image using Sauvola's thresholding (e) Segmented image using NICK's thresholding

Table 3.1 Accuracy of the thresholding techniques

| Sl. no. | Thresholding technique | Accuracy (%) |
|---|---|---|
| 1 | Bernsen's thresholding | 93.6 |
| 2 | Niblack's thresholding | 96 |
| 3 | Sauvola's thresholding | 95 |
| 4 | NICK's thresholding | 95.4 |

Table 3.2 Computation time of different thresholding techniques

| Sl. no. | Thresholding technique | Computation time (sec.) |
|---|---|---|
| 1 | Bernsen's thresholding | 3.10 |
| 2 | Niblack's thresholding | 3.27 |
| 3 | Sauvola's thresholding | 2.74 |
| 4 | NICK's thresholding | 1.73 |

# 4

# Classification of Red Blood Cell and Sickle Cell Using Naïve Bayes Classifier and K-Nearest Neighbor Classifier

## 4.1 Introduction

In this chapter, application of supervised classifiers such as Naïve Bayes classifier as well as K-nearest neighbor classifier are described for sickle cells detection along with a hybrid

segmentation technique. The novel hybrid segmentation technique is proposed for normal and abnormal cell segmentation by combining fuzzy C-means (FCM) and adaptive thresholding method. FCM is combined with four different thresholding methods. From the hybrid segmentation technique, 8 shapes or geometrical features as mentioned below are extracted from each of the sample cells present in the microscopic image of the database.

1. Major axis
2. Minor axis
3. Aspect ratio
4. Area
5. Perimeter
6. Form factor (metric value)
7. Eccentricity and
8. Solidity

These features are fed to both of the classifiers and performance were evaluated and compared for each of the hybrid combination of segmentation using accuracy, sensitivity, specificity and precision. In terms of these performance parameters, the comparison clearly depicts that the proposed hybrid method including fuzzy C-means and NICK's thresholding in combination with K-nearest neighbor classifier performs the best with a significant accuracy of 98.87% along with the lowest time of computation.

In current research scenario of SCA detection using automatic classification techniques, many researchers have evaluated favorable and significant results by using faster and robust systems and various available image processing methods. It is always interesting and encouraging to obtain fresh findings through application of new methodologies for detecting

abnormal RBCs. The motive of this proposed methodology is to formulate a robust and fast segmentation method of hybrid nature for SCA's automatic detection. This work shows combining of adaptive thresholding methods like Bernsen, Niblack, Sauvola and NICK with fuzzy c-means clustering for classifying normal and abnormal RBCs and the performances are then evaluated by use of the some performance parameters for the Naive Bayes and KNN classifiers. The formulation of this hybrid proposed methodology is a novel approach for identification of SCA.

## 4.2 System Block Diagram

The methodology of the proposed work can be well divided into four distinct steps:

1. Image pre-processing,
2. Hybrid image segmentation applying fuzzy C-means and adaptive thresholding methods,
3. Feature extraction and
4. Classification.

The system can be represented in the form of a block diagram as shown in Fig.4.1 [75]. This work is carried out on images collected from online image library [89]. The dataset consists of ten sample microscopic blood images of sickle cell anaemic persons. These images includes both normal as well as abnormal (sickle shaped) RBCs which are fed to the first step of the proposed system as an input for onward processing.

### 4.2.1 Image pre-processing

For extracting fruitful results in any image processing applications, pre-processing of the input images carry significant importance. In this phase the input images are processed for

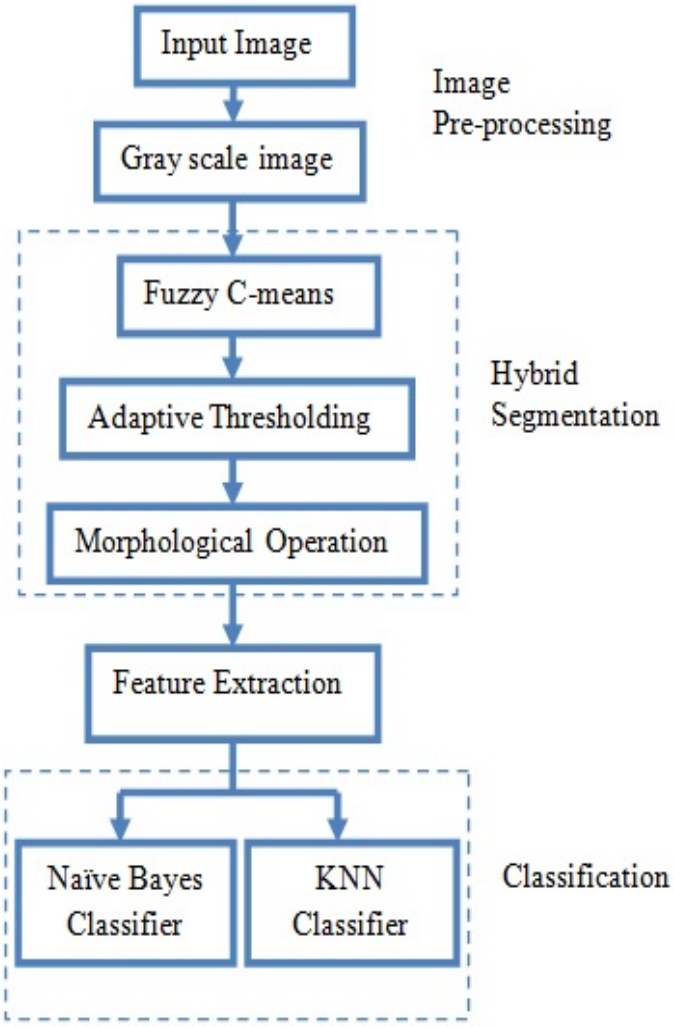


Fig. 4.1 System block diagram of the proposed system

improving its overall quality for ease of application of various image processing techniques effectively. Filtering, debluring, histogram equalization, contrast enhancement etc. are several practises that exist for pre-processing of images. Here also the sample raw images are transformed to gray scale from RGB for enhancing the quality of information included in the images for their effective processing in rest of the stages of the proposed method. Here the image can be filtered using a median filter for removing the impulsive noise present in it if necessary. This filter is used here as it is very useful for preservtion of edges while removing noise.

### 4.2.2 Image Segmentation

Image segmentation is one of the most important step of the system. The primary objective of this segmentation is to change the representation of the images for making its analysis easier. Based on the similarities of the properties, segmentation divides the image into various regions. It plays the important role of identifying region of interest (ROI) in the field of medical image processing [96].

In the context of the current work, the purpose of this step is to separate normal (healthy) and abnormal (unhealthy) RBCs contained in the image. In this work, in order to formulate a hybrid method for segmentation, thresholding based segmentation and clustering based segmentation methods are combined. It is a novel approach for identification of SCA. Here as a hybrid approach at first, each of the microscopic blood images are segmented by fuzzy C-means clustering which are then segmented by using adaptive thresholding methods.

In case of global thresholding method, value of the threshold is fixed. The pixels of the images having higher level of intensity compared to the threshold are assumed as black or foreground and the rest of the pixels are considered as white or background. In all types of images this method does not give encouraging results. For images with variable background and diversity all through the object, the performance of the global thresholding techniques is not proper. In these types of images, satisfactory results are obtained in a few regions while the results obtained in the rest of the regions are not appropriate. To nullify this drawback of global thresholding, local thresholding method is applied. In this technique, the input image is separated into numbers of sub-images. Threshold values for each of the sub-images are then calculated. These calculations of threshold values for the sub-images are done by application of different statistical tools such as mean, standards deviation etc. For the interest of having appropriate results, merging of the threshold values calculated for all the sub-images are done.

In the current work, the images segmented by fuzzy C-means method are processed by four adaptive thresholding techniques namely Niblack, Bernsen, Sauvola and NICK methods.

**Morphological operation**

After performing clustering and thresholding on the input images by using four different methods already mentioned, some morphological operations are to be applied for getting the final segmented images. The purpose of applying these operations is to remove distortions from the output images of the thresholding step. Binary images are used to have many internal holes that are filled. At the next step, the borders of the images are cleared by removing the incomplete items. In addition to these there are some unwanted objects of small sizes in the image that are also eliminated from the final images. After successful applications of these morphological operations, the cells of interest are marked as distinct objects. Appropriate features from these objects are then extracted separately.

### 4.2.3 Extraction of features from the segmented images

Before classification of images, selection and extraction of appropriate features is very important. The extracted features of the image should carry enough information of interest to be fed to the classification stage. As SCA is detected based on changing the shape of the RBCs, hence selection of morphological features of the cells in this work is considered to be the appropriate one.

The feature metric value is the measure of roundness of the RBCs. This is also termed as circular shape factor (CSF). For a normal RBC having round shape the value of CSF is 1 but for sickle shaped RBCs, the value decreases from 0.7 to 0.

$$CSF \quad or \quad Metric \quad value = \frac{4 \times \ pi \times Area}{perimeter^2} \tag{4.1}$$

The ratio of the length of the major axis of the RBC and that of the minor axis is called the Aspect Ratio. For a healthy RBC the value of the ratio is approximately equal to 1 or slightly greater than 1. But for a sickle shaped RBC having elongated shape, its value is much larger than 1 [62].

$$Aspect \quad ratio = \frac{Major \quad axis \quad length}{Minor \quad axis \quad length} \tag{4.2}$$

The features eccentricity and solidity are also evaluated for each of the RBCs of the blood images. The value of eccentricity gives how much the shape of the cells differs from the circular shape. For the round shape, value of eccentricity is 0. The solidity can be evaluated by equation:

$$Solidity = \frac{Area}{Convex \quad hull \quad area} \tag{4.3}$$

For all the healthy and abnormal cells of the segmented blood images the above mentioned geometrical features are extracted. As four different thresholding techniques were applied for hybrid segmentation of the images, hence four distinct sets of features are obtained for each of the RBCs present in the images. These values of the features are then considered for classification.

### 4.2.4 Classification using Naïve Bayes and KNN classifiers

Like feature extraction, selection of appropriate classification techniques also plays an important role in the performances of any proposed methodology. From the literatures reviewed in the context of classification of SCA, classifiers namely Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Artificial Neural Network (ANN) are found to have significant results [43, 44, 53, 54]. Out of these available techniques, for the reported

work Naïve Bayes and KNN classifiers are selected to be used for differentiating the two classes of RBCs.

Table 4.1 Extracted features of a sample image using hybrid segmentation method

| Sl. No. | Area | Perimeter | Metric Value | Major Axis | Minor Axis | Aspect Ratio | Eccentricity | Solidity |
|---|---|---|---|---|---|---|---|---|
| Normal RBC1 | 270 | 57.39 | 1.02 | 20.17 | 17.61 | 1.14 | 0.48 | 0.97 |
| Normal RBC2 | 654 | 92.33 | 0.96 | 32.02 | 26.92 | 1.18 | 0.54 | 0.95 |
| Normal RBC3 | 409 | 69.27 | 1.07 | 23.59 | 22.16 | 1.06 | 0.34 | 0.97 |
| Normal RBC4 | 431 | 71.19 | 1.06 | 23.79 | 23.13 | 1.02 | 0.23 | 0.97 |
| Normal RBC5 | 437 | 71.43 | 1.07 | 25.12 | 22.19 | 1.13 | 0.46 | 0.98 |
| Sickle RBC1 | 708 | 153.08 | 0.37 | 52.78 | 23.14 | 2.28 | 0.89 | 0.75 |
| Sickle RBC2 | 712 | 116.04 | 0.66 | 48.16 | 21.90 | 2.19 | 0.89 | 0.84 |
| Sickle RBC3 | 756 | 122.52 | 0.63 | 52.24 | 20.64 | 2.53 | 0.91 | 0.85 |
| Sickle RBC4 | 611 | 125.83 | 0.48 | 55.81 | 18.78 | 2.97 | 0.94 | 0.76 |
| Sickle RBC5 | 974 | 139.03 | 0.63 | 59.93 | 22.19 | 2.70 | 0.92 | 0.88 |

## 4.3 Results and Discussion

The results evaluated in each of the stages of the proposed system are presented in this section with associated analysis. In the pre-processing step, the raw images were transformed to gray scale from RGB for enhancement of its quality which are then fed for segmentation. In the segmentation step, four combinations of hybrid segmentation methods were used. The segmented images obtained from the sample microscopic blood image by using fuzzy C-means pairing with Bernsen, Niblack, Sauvola and NICK thresholding methods respectively are shown in Figure 4.2. The process was repeated for all the images of the dataset. For all the healthy and abnormal cells of the segmented blood images the selected geometrical features are then extracted and tabulated. Some sets of these extracted feature values for five numbers of healthy and sickle cells of one input image are shown in Table 7.1. Each of these feature sets contains eight geometrical parameters of the RBCs. These eight numbers of the geometrical parameters are evaluated for each of the RBCs present in all the ten images taken in the sample dataset by applying four types of hybrid methodologies considered. From the results shown in Table 7.1, it can be concluded that for the healthy RBCs, the ranges of the aspect ratio and metric value are $[1-1.2]$ and $[0.85-1.07]$ respectively. But in case of the sickle RBCs the ranges of the aspect ratio and metric value are $[2-3]$ and $[0.35-0.80]$ respectively. The eccentricity for the healthy RBCs is less than 0.5 whereas for the sickle RBCs this value is greater than 0.5. The solidity on the other hand for the healthy RBCs is equal to 1 approximately, whereas in case of the sickle RBCs, the said value is less than 0.75.

After completion of the evaluation of selected features from the healthy and sickle RBCs of the input images, these are fed to the classifiers Naïve Bayes and KNN for appropriate classification. 60% out of the total extracted feature set of each of the RBCs are used for training the classifiers and the rest of the 40% feature set are considered for testing purpose.

Fig. 4.2 (a) Sample image (b) Result of hybrid segmentation using Bernsen's thresholding (c) Result of hybrid segmentation using Niblack's thresholding (d) Result of hybrid segmentation using Sauvola's thresholding (e) Result of hybrid segmentation using NICK's thresholding

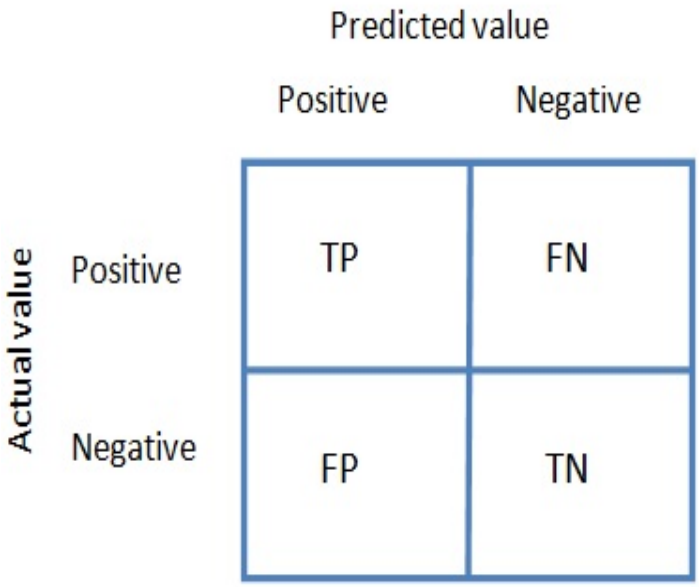


Fig. 4.3 Confusion matrix for calculation of performance parameters

The performance of the classifiers is evaluated using confusion matrix. Using confusion matrix, the parameters for performance of the classifiers such as accuracy, sensitivity, specificity and precision are calculated. These parameters can be explained from confusion matrix shown in Fig. 4.3 as:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{4.4}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{4.5}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.6}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.7}$$

where, TP =True Positive = Quantity of positive class (sickle cell) correctly identified as positive in numbers.

TN =True Negative = Quantity of negative class (normal RBC) properly detected as negative

in numbers.

FP =False Positive = Quantity of negative class (normal RBC) inaccurately detected as positive (sickle cell) in numbers.

FN =False Negative= Quantity of positive class (sickle cell) incorrectly detected as negative (normal RBC) in numbers.

Accuracy is the ratio of number of cells that are perfectly classified to total number of cells classified by the classifier. Sensitivity is the ratio of perfectly detected unhealthy cells i.e. sickle cells to total unhealthy cells. Specificity is obtained by dividing accurately detected healthy cell to total healthy cells. Precision is ratio of total accurately detected unhealthy cells and the total predicted unhealthy cells. Using the values of the confusion matrix, four performance parameters of Naïve Bayes and KNN classifiers are calculated for all the images present in the dataset.

Table 4.2 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Bernsen thresholding

| Sl. No. | Naïve Bayes | | | | KNN | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 98.78 | 97.98 | 100 | 100 | 98.75 | 100 | 98.80 | 98.75 |
| Image2 | 96.80 | 96.22 | 98.67 | 98.42 | 96.40 | 97.53 | 96.55 | 96.70 |
| Image3 | 97.58 | 98.54 | 97.62 | 95.63 | 97.52 | 98.69 | 97.51 | 97.61 |
| Image4 | 96.45 | 96.70 | 94.50 | 95.30 | 96.30 | 96 | 95.80 | 96.33 |
| Image5 | 98.28 | 95.19 | 97.10 | 97.56 | 98.80 | 97.46 | 96.39 | 96.84 |
| Image6 | 97.80 | 97.34 | 97.70 | 97.21 | 96.49 | 95.83 | 97.49 | 97.95 |
| Image7 | 95.41 | 96 | 95.70 | 94.67 | 96.91 | 95.28 | 97.36 | 98.53 |
| Image8 | 97.80 | 98.52 | 96.89 | 97.89 | 98.36 | 98.39 | 98.28 | 96.21 |
| Image9 | 96.93 | 97.84 | 97.21 | 97 | 98.80 | 97.62 | 98.40 | 95.83 |
| Image10 | 96.76 | 97.26 | 97.20 | 97.18 | 97.60 | 97.20 | 97.93 | 96.78 |
| Overall | 97.25 | 97.15 | 97.25 | 97.08 | 97.59 | 97.40 | 97.45 | 97.15 |

Table 4.3 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Sauvola thresholding

| Sl. No. | Naïve Bayes | | | | KNN | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 100 | 98.38 | 97.57 | 100 | 98.53 | 97.8 | 100 | 100 |
| Image2 | 97.80 | 97.15 | 96.48 | 96.89 | 97.43 | 97.53 | 97.56 | 96.70 |
| Image3 | 95.32 | 94.69 | 94.79 | 95.78 | 96.20 | 95.66 | 94.67 | 95.66 |
| Image4 | 96.90 | 95.80 | 94.50 | 95.16 | 96.48 | 97.23 | 97.45 | 96.47 |
| Image5 | 97.40 | 98.27 | 97.36 | 97.40 | 97.67 | 97.44 | 98.12 | 97.57 |
| Image6 | 96.80 | 97.30 | 98.56 | 97.45 | 97.37 | 97.80 | 97.33 | 96.78 |
| Image7 | 97.12 | 97.28 | 97.67 | 97.80 | 98.46 | 98.23 | 98.60 | 97.33 |
| Image8 | 97.80 | 97.80 | 96.78 | 96.34 | 97.76 | 97.22 | 97.31 | 97.77 |
| Image9 | 95.80 | 96.90 | 97.67 | 96.20 | 97.55 | 97.62 | 96.78 | 97.38 |
| Image10 | 98.34 | 98.57 | 100 | 100 | 98.42 | 97.56 | 100 | 100 |
| Overall | 97.32 | 97.21 | 97.13 | 97.30 | 97.58 | 97.40 | 97.78 | 97.56 |

Table 4.4 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using Niblack thresholding

| Sl. No. | Naïve Bayes | | | | KNN | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 100 | 100 | 100 | 100 | 97.88 | 98.10 | 97 | 97.33 |
| Image2 | 97.55 | 96.88 | 95.57 | 94.78 | 97.40 | 96.68 | 96 | 96.44 |
| Image3 | 98.41 | 97.58 | 97.45 | 96.68 | 98.82 | 98.33 | 97.44 | 96.80 |
| Image4 | 94.78 | 95.54 | 95.51 | 95 | 95.78 | 94.48 | 96.60 | 96.22 |
| Image5 | 97.66 | 97.35 | 97.64 | 96.78 | 100 | 100 | 100 | 100 |
| Image6 | 98.11 | 98 | 97.61 | 97.33 | 97.80 | 97.80 | 98 | 98.21 |
| Image7 | 96.66 | 95.59 | 97.77 | 97.11 | 95.55 | 96.22 | 96.68 | 97.30 |
| Image8 | 96.98 | 96.43 | 96.22 | 97 | 97.62 | 96.66 | 97.12 | 97.30 |
| Image9 | 97.52 | 97.79 | 98.90 | 97.80 | 98.83 | 98 | 98.55 | 98.58 |
| Image10 | 98.56 | 98.17 | 98.80 | 97.89 | 98.44 | 97.90 | 97.11 | 96.80 |
| Overall | 97.62 | 97.33 | 97.54 | 97.03 | 97.81 | 97.41 | 97.45 | 97.49 |

Table 4.5 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using NICK thresholding

| Sl. No. | Naïve Bayes | | | | KNN | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 98.46 | 98.22 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image2 | 98.30 | 98.48 | 100 | 100 | 98.41 | 98.40 | 98.25 | 97.89 |
| Image3 | 97.87 | 97.67 | 98.54 | 96.70 | 100 | 98.58 | 98.76 | 98.64 |
| Image4 | 97.41 | 97.58 | 95.70 | 97.12 | 98.42 | 97.54 | 97.28 | 97.39 |
| Image5 | 100 | 98.73 | 97.39 | 97.56 | 100 | 100 | 100 | 100 |
| Image6 | 98.60 | 97.29 | 98.56 | 98.83 | 97.28 | 97.38 | 96.48 | 96.39 |
| Image7 | 97.80 | 98.90 | 95.70 | 94.67 | 97.60 | 98.65 | 98.40 | 98.61 |
| Image8 | 98.59 | 97.64 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image9 | 97.40 | 97.69 | 98.50 | 98.32 | 98.48 | 96.68 | 98.48 | 97.2 |
| Image10 | 97.69 | 98.18 | 97.80 | 97.43 | 98.59 | 98 | 97.30 | 97.57 |
| Overall | 98.21 | 98.03 | 98.21 | 98.06 | 98.87 | 98.52 | 98.49 | 98.36 |

Table 4.6 Overall performance of Naïve Bayes and KNN classifier with hybrid segmentation using four thresholding methods

| Thresholding method | Naïve Bayes | | | | KNN | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Bernsen | 97.25 | 97.15 | 97.25 | 97.08 | 97.59 | 97.40 | 97.45 | 97.15 |
| Sauvola | 97.32 | 97.21 | 97.13 | 97.30 | 97.58 | 97.40 | 97.78 | 97.56 |
| Niblack | 97.62 | 97.33 | 97.54 | 97.03 | 97.81 | 97.41 | 97.45 | 97.49 |
| NICK | 98.21 | 98.03 | 98.21 | 98.06 | 98.87 | 98.52 | 98.49 | 98.36 |

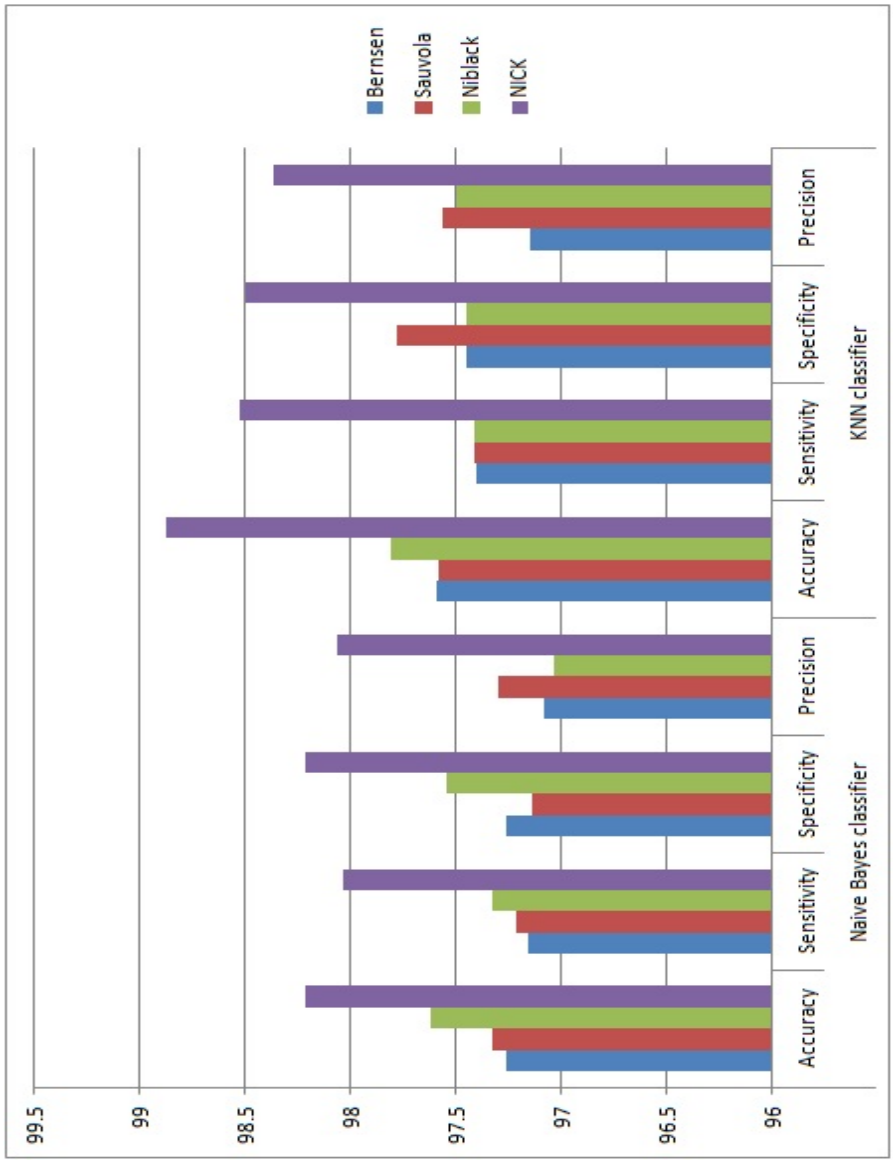


Fig. 4.4 Performance of Naïve Bayes and KNN classifier with hybrid segmentation using four thresholding methods

Table 4.7 Computation time of Naïve Bayes and KNN classifier

| Segmentation method | Computation time (sec.) | |
|---|---|---|
| | Naïve Bayes | KNN |
| Fuzzy C-means and Bernsen Thresholding | 0.88 | 0.79 |
| Fuzzy C-means and Sauvola Thresholding | 0.86 | 0.81 |
| Fuzzy C-means and Niblack Thresholding | 0.82 | 0.76 |
| Fuzzy C-means and NICK Thresholding | 0.56 | 0.45 |

The performances obtained from the results of the classifiers Naïve Bayes and KNN using the thresholding methods Bernsen, Sauvola, Niblack and NICK respectively are shown in Table 4.2 - Table 4.5. Table 4.6 presents the overall values of the parameters accuracy, sensitivity, specificity and precision for the applied classifiers for all four types of adaptive thresholding techniques. The graphical representation of these values of the performance parameters are represented in Figure 4.4. The comparison presented in these results shows that the hybrid segmentation technique using the four adaptive thresholding methods gives satisfactory results in detection of SCA. Amongst the four adaptive thresholding techniques, NICK's method gives better performance in case of both Naïve Bayes and KNN classifiers. It can also be seen in the results that, the performance of KNN classifier is comparatively superior to that of the Naïve Bayes classifier. KNN classifier gives highest accuracy of 98.87% with NICK's thresholding method. Table 4.7 shows the computational time of Naïve Bayes and KNN classifier for different thresholding methods. From computational time point of view also, KNN classifier and the hybrid combination of fuzzy C-means and NICK thresholding performs better.

Finally a comparative analysis of the proposed method with other existing methods based on the non-overlapping sickle cell detection is carried out and is presented in Table 4.8. From this comparison, it is seen that the performance of this proposed method using KNN classifier with fuzzy C-means and NICK thresholding technique is the best.

Table 4.8 Comparison of performance of our method with other existing methods

| Method | Performance measures | | | |
|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision |
| Proposed method (KNN classifier with fuzzy C-means and NICK thresholding) | 98.87 | 98.52 | 98.49 | 98.36 |
| Aliyu *et al.* [17] | 93 | 94 | 80 | - |
| Rakshit *et al.* [30] | 95.8 | - | - | - |
| Tomari *et al.* [43] | 83 | 76 | - | 82 |
| Chy *et al.* [50] | 95 | 96.55 | - | - |

# 5

# Classification of Sickle Cell based on Adaptive Thresholding and Contour-based methods using ANN and SVM Classifiers

## 5.1 Introduction

The works carried out with the motive of SCA's automatic detection by use of a novel segmentation technique in combination with four different methods of thresolding and active contour-based algorithm are included in this chapter.

In this work, for the purpose of detection of sickle cells, supervised classifiers like Artificial Neural Network (ANN) and Support Vector Machine (SVM) are selected. Amongst the four selected adaptive thresholding techniques the performance of NICK's methods is found to be superior than the others. Again the performance of the SVM classifier is found to be slighly better than ANN classifier while trained with scaled conjugate gradient back-propagation algorithm with 10 hidden layer neurons. This proposed approach results in highest accuracy of 99.2% using SVM classifier along with NICK's thresholding method and active contour-based algorithm. The performance of the reported approach is also analysed for seven different types of training algorithms in the ANN classifiers by changing the numbers of hidden layer neurons. On comparing the performances of these algorithms, it was seen that resilient back-propagation algorithm with 10 hidden layer neurons gave resonably better performance in ANN with the accuracy of 99%. The evaluated results from this approach and its analysis proves that ANN and SVM classifiers used in combination with adaptive thresholding methods and active contour based technique is an effective and fruitful approach of classification of SCA patient. Over the past years, various segmentation techniques for the analysis of medical images have been reported by researchers. Region based segmentation methods were amongst such techniques which were widely used by the researchers. Some of these region based detection methods implemented on medical images are region growing, contour based technique and region based level set technique etc. [28, 98, 99]

In recent years, significant increase in the application of neural network in processing of medical images for the purpose of classification has been noticed because of its high

accuracy and less computational time [100, 101]. Khalaf *et al.*[45] reported an approach of using different types of neural network models for sickle cell disease detection from the considered biomedical dataset. Another algorithm was proposed by Elsalamony [44] for detection of anaemia by neural network. In this algorithm identification of the abnormal RBCs from microscopic blood images was done by using CHT and morphological operations. In a similar approach, 98%, 100%, 100%, and 99.3% rates of effectiveness for detecting elliptocytosis RBCs, sickle cells, microsite RBCs and unknown shaped RBCs respectively were achieved by the authors from the 21 numbers of microscopic blood images using artificial neural networks. The support vector machine classifier on the other hand was trained for SCA's automatic detection [50] which gave 95% and 96.55% of accuracy and sensitivity respectively. The use of different machine learning methodologies such as ANN trained with random forest classifier, Levenberg-Marquardt algorithm, Elman & Jordan recurrent neural network (RNN) classifier, hybrid RNN, combining the Elman and Jordan networks and SVM classifier for detection of SCD was reported in[41].

## 5.2 System Block Diagram

The proposed methodology in the form of a system block diagram is presented in Fig. 5.1. The methodology can be well described with the help of five major blocks as follows:

1. Input image data
2. Pre-processing of the input images
3. Image segmentation by Active contour-based technique and four adaptive thresholding methods with morphological operations
4. Feature extraction and
5. Classification by using ANN and SVM classifiers [78]

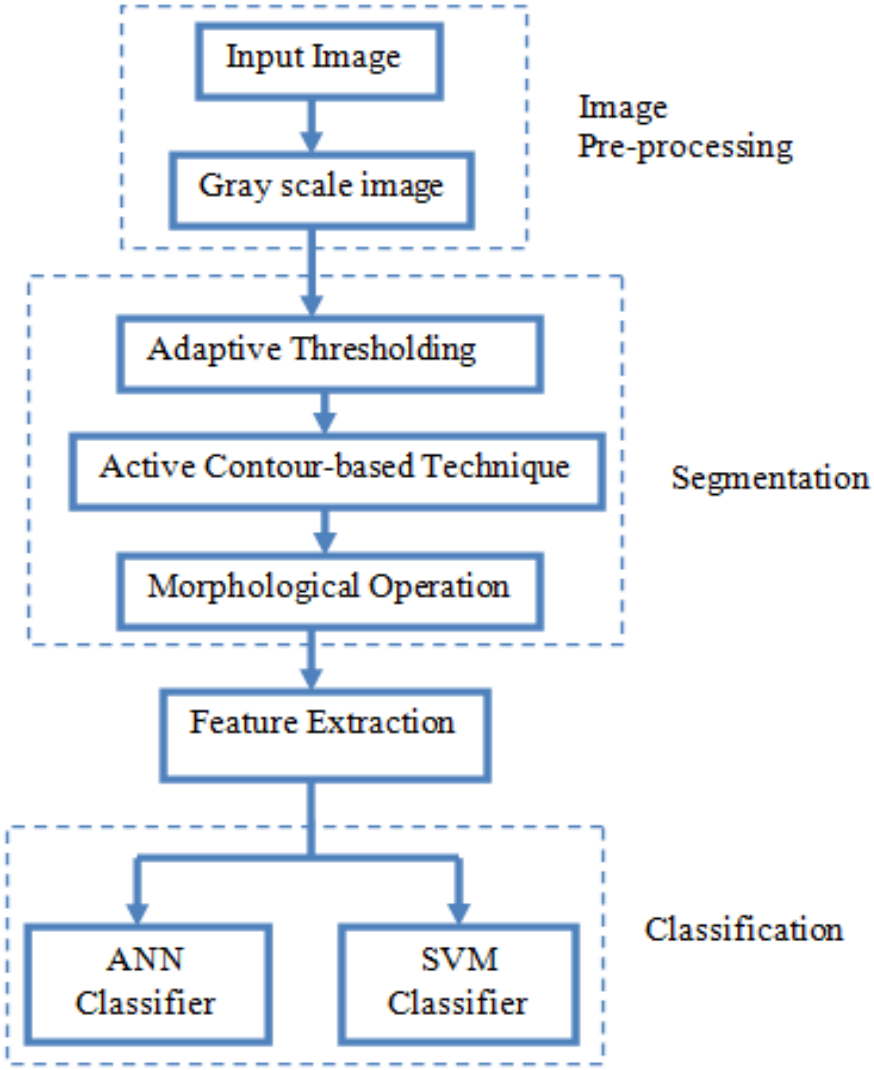


Fig. 5.1 System block diagram

### 5.2.1 Pre-processing of input data sample

This work is carried out on images collected from online image library [89]. The dataset consists of ten sample microscopic blood images of sickle cell anaemic persons. These images includes both normal as well as abnormal (sickle shaped) RBCs which are fed to the first step of the proposed system as an input for onward processing. The images are transformed to gray scale from RGB for enhancing the quality of information included in the images for their effective processing in rest of the stages of the proposed method. The

prime objective of this pre-processing is to prepare the input data or images in an appropriate manner so that significant and fruitful results can be evaluated in its processing steps as desired.

### 5.2.2 Segmentation of pre-processed images

Image thresholding is considered to be the simplest type of image segmentation process, as it segregates the image into two regions of pixels, i.e. foreground and background. The main objective of medical image segmentation is to identify the region of interest from these two based on the type of application. In this work, the normal and abnormal RBCs are divided from the background during segmentation. For this purpose, four local adaptive thresholding methods along with active contour-based segmentation techniques are used on the pre-processed blood images.

The Active contour model is a region based iterative boundary detection method starting with an initial contour for target boundary. The contour is then changed so that it reaches the final desired boundary of the object with some defined criteria.

In case of local adaptive thresholding, for the sub-images of the whole image more than one threshold values are considered. Some characteristics of the local image areas are utilised to select a different threshold values for distinct parts of the image. Different adaptive thresholding methods with significant performance capabilities are used for biomedical images segmentation which includes Niblack's, Sauvola's, Bernsen's and NICK's thresholding. In Niblack thresholding, the algorithm computes the threshold in accordance with the local mean and standard deviation over a particular size of window throughout each pixel position in the image [102]. The governing equation of NICK's thresholding is derived from the Niblack's thresholding [95]. In this work, four thresholding methods are experimented with active contour-based segmentation on the pre-processed input images followed by some selected morphological operations in order to obtain the final segmented images. The spe-

cific objective of performing morphological operation in the above process is to nullify the unwanted distortions or incomplete objects from the images already thresholded. The images in the database are also segmented by using a global thresholding method namely Otsu's thresholding along with active contour-based method for comparing the performances of adaptive and global thresholding. Figure 5.2 shows the results of hybrid segmentation with Otsu's global thresholding and four adaptive thresholding methods respectively.

### 5.2.3 Feature extraction from segmented images

In most of the cases morphological features are used to classify normal and abnormal RBCs for recognition of sickle cell disease. Seven geometrical features are extracted from each detected cell of the sample image. These are area, perimeter, major axis, minor axis, circularity, eccentricity and solidity. Using the values of the major axis and minor axis thus extracted, aspect ratio is evaluated as given by the Eq. 5.1.

$$Aspect \quad Ratio = \frac{Major \quad Axis \quad Length}{Minor \quad Axis \quad Length} \tag{5.1}$$

For the RBCs with normal shape, the value of the aspect ratio is nearly equal to 1, whereas because of the crescent shape of sickle cell, it's value is much greater than 1. The Circularity or metric value or effect factor can be calcuted from the features area and perimeter using the Eq. 5.2.

$$Circularity \quad or \quad Metric \quad Value = \frac{4 \times \ pi \times area}{perimeter^2} \tag{5.2}$$

The value of Circularity is 1 for normal RBCs and approximately less than 0.7 for sickled RBCs. The degree of deviation of an object from being circular can be measured by Eccentricity. It's value for normal RBCs is 0 for being round in shape.

$$Solidity = \frac{Area}{Area \quad of \quad Convex \quad hull} \tag{5.3}$$

The value of Solidity is 1 and less than 1 for an object with regular boundary and irregular boundary respectively. Four different sets of features for each of the image in dataset are obtained after applying thresholding methods namely Bernsen, Niblack, Sauvola and NICK's thresholding. Figure 5.3 shows the final segmented image of one cropped sample image with 8 feature values.

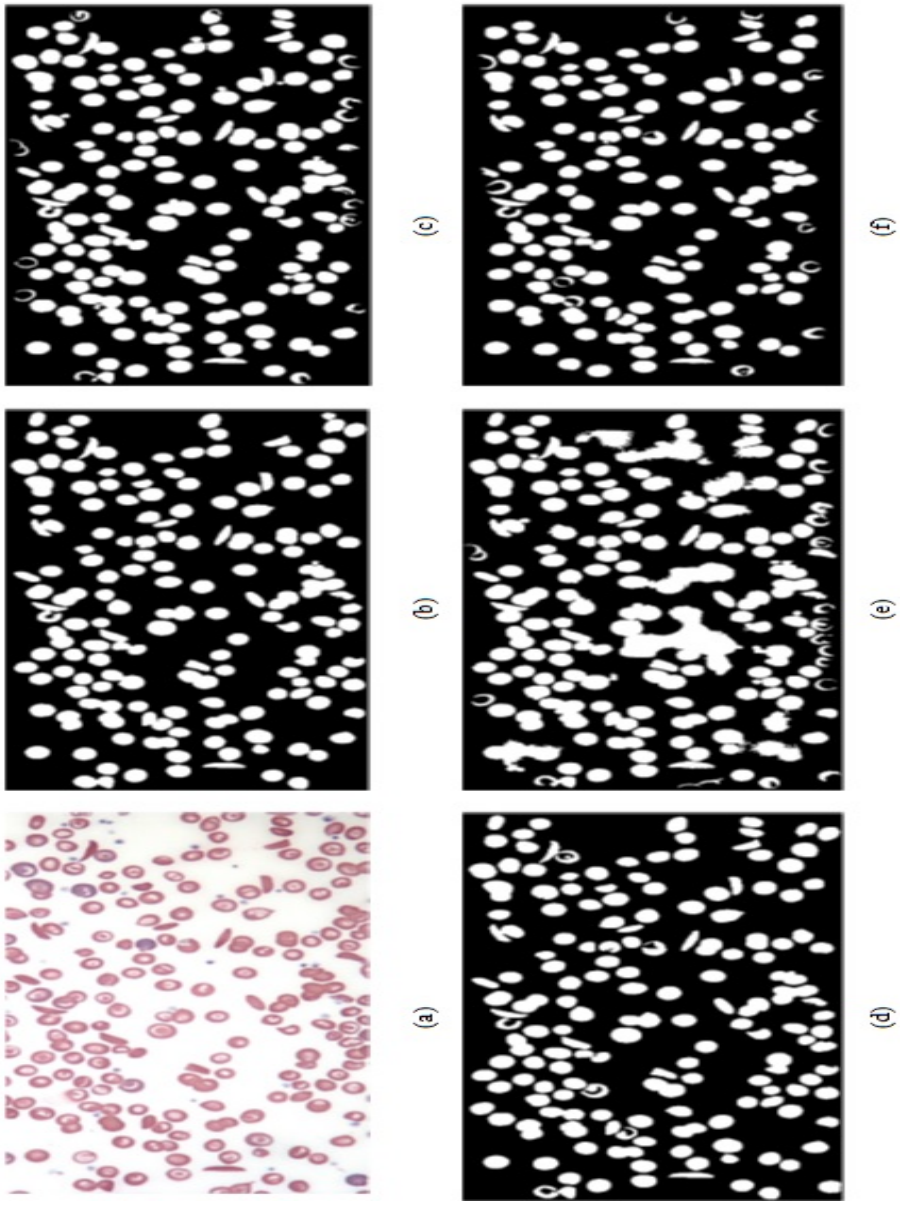


Fig. 5.2 (a) Sample image (b) Result of hybrid segmentation using Otsu's global thresholding (c) Result of hybrid segmentation using Bernsen's thresholding (d) Result of hybrid segmentation using Niblack's thresholding (e) Result of hybrid segmentation using Sauvola's thresholding (f) Result of hybrid segmentation using NICK's thresholding

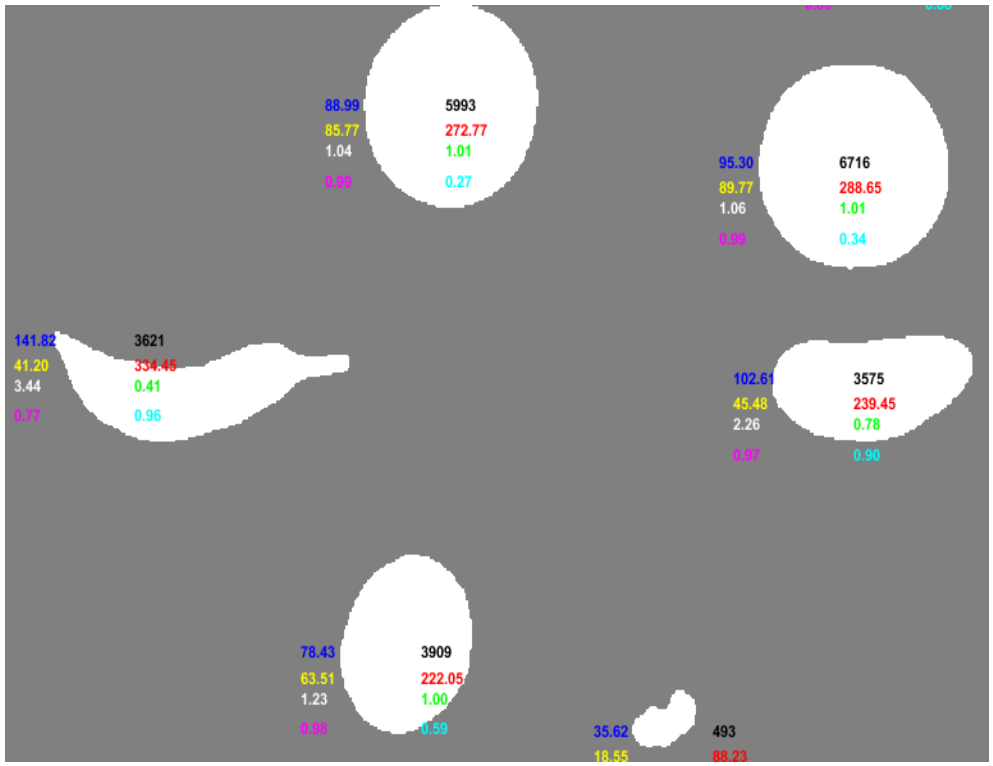


Fig. 5.3 Extracted features of normal and sickle RBCs from a cropped sample image [Blue colour=Major axis, Yellow colour=Minor axis, , White color=Aspect Ratio, Magenta colour=Solidity, Black colour=Area, Red colour=Perimeter, Green color=Metric value , Cyan color=Eccentricity]

Table 5.1 Feature values for a normal and sickle cell

| RBC type | Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Major Axis | Minor Axis | Aspect ratio | Solidity | Area | Perimeter | Circularity | Eccentri-city |
| Normal | 87.46 | 84.37 | 1.03 | 0.99 | 5938 | 273.28 | 1 | 0.27 |
| Sickle | 104.58 | 43.87 | 2.38 | 0.97 | 3496 | 241.05 | 0.76 | 0.90 |

### 5.2.4 Classification of images using ANN and SVM classifiers

The feature vectors for different classifiers are formed from the features extracted from all the normal and abnormal cells for each of the images considered in the dataset. Such a feature vector for a normal as well as sickle cell is shown in Table 5.1. The table clearly depicts the differences in values of the extracted features for normal and sickle RBC. Thus the feature vectors are derived for all normal and sickle cells for 10 images of the dataset from segmentation method and are applied for training the ANN and SVM classifiers one by one. Then the performance measures of both the classifiers are studied. In this process different training algorithms were used for training in the ANN and the evaluatied results were also studied in a comparative manner. The back-propagation (BP) algorithms considered here are: Scaled conjugate gradient BP (SCGBP), Gradient descent with adaptive learning rate BP (GDALBP), Levenberg-Marquardt BP (LMBP), Gradient descent with momentum BP (GDMBP), Bayesian regularization BP (BRBP), Gradient descent with momentum and adaptive learning rate BP (GDMALBP) and resilient BP (RBP).

## 5.3 Results and Discussion

In the current work, machine learning techniques ANN with BP algorithm and SVM are used for classification of normal and sickle cell. In case of ANN, 8 input layer features, 2 output layer neurons and 10 numbers of intermediate layer neurons are used along with

scaled conjugate gradient BP algorithm for training and testing purpose. Whole the input data is divided into 3 sections. 70% of that are used for training, 15% are used for validation and the remaining 15% data are used for testing purpose. In the context of SVM classifier, 70% and 30% samples are used for training and testing purpose respectively.

For performance analysis of the ANN and SVM classifiers, we evaluated the percentage of correct classification rate or accuracy, sensitivity, specificity and precision for each of the classifier using confusion matrix. Figure 5.4 shows the confusion matrix of ANN classifier with active contour-based method and NICK thresholding for a sample image. Table 5.2- Table 5.5 show the performance parameters of ANN and SVM classifier using Bernsen, Sauvola, Niblack and NICK's thresholding along with active contour segmentation respectively. Table 5.6 shows the overall performance of ANN and SVM classifiers for the dataset. Amongst the four adaptive thresholding techniques, NICK's method gives better performance in case of both the ANN and SVM classifiers. In addition the analysis also shows that the performance of SVM classifier is better over the ANN type for sickle cell detection. The proposed segmentation technique with NICK thresholding results in maximum accuracies of 99.2% and 98.8% for SVM and ANN classifier respectively with scaled conjugate gradient back-propagation algorithm. The system is also experimented by using global thresholding method namely Otsu's thresholding in place of adaptive thresholding and results are shown in Table 5.7. Comparing the results of Table 5.6 and Table 5.7, it is found that the performance of adaptive thresholding is better than Otsu's global thresholding.

At the next stage, with different training algorithms the performance of the ANN classifier is evaluated by varying the numbers of neurons in the hidden layer. The overall accuracies obtained in case of 7 different training algorithms used to train the ANN classifier in this work are shown in Table 5.8. Out of these back-propagation algorithms, Resilient back-propagation algorithm shows the highest accuracy of 99% with 10 numbers of hidden layer neurons.

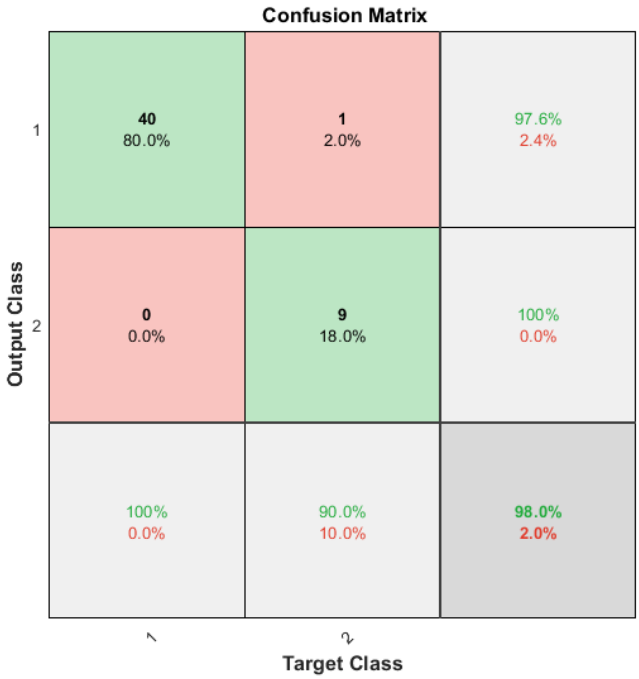


Fig. 5.4 Confusion matrix of ANN classifier

Table 5.2 Performance of ANN and SVM classifier using Bernsen thresholding and active contour-based segmentation

| Sl. No. | ANN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 96.4 | 98 | 96 | 94.6 | 97.6 | 100 | 97 | 96.4 |
| Image2 | 98.6 | 97.4 | 100 | 100 | 98 | 97.3 | 100 | 100 |
| Image3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image4 | 97.6 | 96.8 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image5 | 88.4 | 100 | 87.8 | 84.7 | 91.7 | 100 | 90.8 | 89.6 |
| Image6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image8 | 98.8 | 97.8 | 98.4 | 94.7 | 96 | 94.7 | 100 | 100 |
| Image9 | 98.2 | 100 | 98.6 | 97.6 | 98.2 | 100 | 97.8 | 95.8 |
| Image10 | 98.8 | 97.8 | 100 | 100 | 97.4 | 96.8 | 100 | 100 |
| Overall | 97.6 | 98.8 | 98 | 97.2 | 97.8 | 98.8 | 98.5 | 98.1 |

Table 5.3 Performance of ANN and SVM classifier using Sauvola thresholding and active contour-based segmentation

| Sl. No. | ANN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 100 | 100 | 100 | 100 | 98.6 | 95.2 | 100 | 100 |
| Image2 | 95.3 | 93.5 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image3 | 96.7 | 97.5 | 95.4 | 92.1 | 100 | 100 | 100 | 100 |
| Image4 | 100 | 100 | 100 | 100 | 92.3 | 100 | 89 | 84.7 |
| Image5 | 97.4 | 100 | 95.2 | 92.5 | 96.8 | 93.2 | 100 | 100 |
| Image6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image8 | 92.3 | 100 | 88.7 | 83.6 | 98.4 | 93.5 | 100 | 100 |
| Image9 | 100 | 100 | 100 | 100 | 92 | 100 | 84.6 | 82.1 |
| Image10 | 94.7 | 92.6 | 100 | 100 | 100 | 100 | 100 | 100 |
| Overall | 97.6 | 98.3 | 97.9 | 96.8 | 97.8 | 98.1 | 97.3 | 96.6 |

Table 5.4 Performance of ANN and SVM classifier using Niblack thresholding and active contour-based segmentation

| Sl. No. | ANN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image2 | 94.5 | 93.5 | 95.7 | 92.2 | 97.4 | 93.6 | 100 | 100 |
| Image3 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image4 | 98.6 | 96.8 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image5 | 94.6 | 100 | 92.4 | 90.6 | 92.6 | 100 | 89.4 | 86.8 |
| Image6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image8 | 97.2 | 97.8 | 96.2 | 93.6 | 97.2 | 94.7 | 100 | 100 |
| Image9 | 95.7 | 100 | 98.2 | 96.8 | 95.3 | 100 | 93.4 | 91.7 |
| Image10 | 96.4 | 97.8 | 100 | 100 | 98 | 96.8 | 100 | 100 |
| Overall | 97.7 | 98.5 | 98.2 | 97.3 | 98 | 98.5 | 98.2 | 97.8 |

Table 5.5 Performance of ANN and SVM classifier using NICK's thresholding and active contour-based segmentation

| Sl. No. | ANN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Image1 | 98.6 | 100 | 98.5 | 95.7 | 100 | 100 | 100 | 100 |
| Image2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image3 | 98 | 100 | 97 | 94.3 | 95.8 | 92.4 | 100 | 100 |
| Image4 | 100 | 100 | 100 | 100 | 98.6 | 100 | 94.5 | 92.3 |
| Image5 | 96.7 | 98 | 96.3 | 92.8 | 100 | 100 | 100 | 100 |
| Image6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Image9 | 97.1 | 96.4 | 98 | 96.3 | 98.5 | 94.6 | 100 | 100 |
| Image10 | 98 | 100 | 96.4 | 94.2 | 98.8 | 100 | 93.4 | 90 |
| Overall | 98.8 | 99.4 | 98.6 | 97.3 | 99.2 | 98.7 | 98.7 | 98.3 |

Table 5.6 Overall performance of ANN and SVM classifier using four adaptive thresholding methods and active contour-based segmentation

| Thresholding method | ANN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Precision | Accuracy | Sensitivity | Specificity | Precision |
| Bernsen | 97.6 | 98.8 | 98 | 97.2 | 97.8 | 98.8 | 98.5 | 98.1 |
| Sauvola | 97.6 | 98.3 | 97.9 | 96.8 | 97.8 | 98.1 | 97.3 | 96.6 |
| Niblack | 97.7 | 98.5 | 98.2 | 97.3 | 98 | 98.5 | 98.2 | 97.8 |
| NICK | 98.8 | 99.4 | 98.6 | 97.3 | 99.2 | 98.7 | 98.7 | 98.3 |

Table 5.7 Overall performance of ANN and SVM classifier using Otsu's thresholding and active contour-based segmentation

| Classifier | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| ANN | 96.8 | 98.7 | 96.1 | 87.9 |
| SVM | 97.4 | 97.9 | 98.9 | 98.8 |

The proposed approach of classification of blood images for detection of SCA is compared with some other existing methods used for SCA detection and the results of comparison are shown in Table 5.9. In this result, performance of SVM classifier with NICK's thresholding and active contour-based segmentation method has been shown in comparison with the existing methods of SCA detection which are based on global thresholding segmenting technique. This comparative analysis clearly shows improved performances of ANN and SVM classifiers while used with local adaptive thresholding method.

This work has been carried out with a distinct motive of making the process of detection of SCA automatic with the help of some appropriate segmentation methods. For this purpose, machine learning techniques like ANN and SVM are applied for effective detection of sickle cells. Again for segmentation, adaptive thresholding techniques with active contour-based methods are used. From the analysis of the evaluated results, it is seen that, both the ANN and SVM classifiers give maximum classification rate with NICK's and active contour based segmentation methods. In terms of accuracies and other considered performance parameters, both these classifiers gave better results in case of adaptive thresholding methods than Otsu's global thresholding technique.

The analyses of the results are further extended by comparing the performances of ANN evaluated with seven different and unique training algorithms. This analysis shows that resilient BP algorithm gave moderately better accuracy than the others. While comparing the performance of the proposed approach of SCA detection with other similar methods, favourable results were also obtained in support of the proposed method. In view of the entire

Table 5.8 Overall Accuracy of ANN classifier with different training algorithms and varying hidden layer neurons

| Sl. No. | ANN training algorithm | No.s of hidden layer neurons | | | |
|---|---|---|---|---|---|
| | | **5** | **10** | **50** | **100** |
| 1 | SCGBP | 98 | 98.8 | 97.3 | 96.8 |
| 2 | GDMBP | 97 | 97.3 | 95.7 | 93.7 |
| 3 | GDALBP | 95.9 | 97 | 94.7 | 93.5 |
| 4 | GDMALBP | 98 | 98.2 | 97.4 | 96.5 |
| 5 | LMBP | 97.8 | 98.5 | 96.3 | 95.8 |
| 6 | BRBP | 96.9 | 97.8 | 97.4 | 96 |
| 7 | RBP | 98.7 | 99 | 98.3 | 97.5 |

Table 5.9 Performance comparison of proposed system with other techniques

| Technique | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Proposed method (SVM classifier with NICK's thresholding and active contour-based segmentation) | 99.2 | 98.7 | 98.7 | 98.3 |
| Albayrak *et al.* [16] | 91.1 | 79 | - | 92.9 |
| Aliyu *et al.* [17] | 93 | 94 | 80 | - |
| Rakshit *et al.* [30] | 95.8 | - | - | - |
| Chy *et al.* [50] | 95 | 96.5 | - | - |

analysis provided in this work it can be concluded that, for both the normal non-overlapping and the sickle RBCs, ANN and SVM classifiers with adaptive thresholding and active contour based technique is an efficient and effective approach for detection of SCA.

# 6

# Sickle Cell Detection from Blood Images of Anaemic Person using Hybrid Segmentation Method and ELM Classifier

## 6.1 Introduction

Various machine learning techniques are used widely for diagnosis and detection of diseases in medical field. Automatic and fast diagnosis of SCD has vital importance for its early

detection and therefore reducing the severity of the disease. Extreme Learning Machine (ELM) is a kind of feed-forward neural network which is used here for classification of SCD from microscopic blood smear. For classification, feature sets are generated using eight geometrical or shape features of normal as well as sickle cells of images considered in the dataset. In this work, a hybrid technique comprising adaptive thresholding and active contour-based method is used for segmenting and extracting features from cells. The hybrid segmentation is experimented with four different adaptive thresholding methods along with contour-based segmentation. A comparison of these four adaptive thresholding methods is also carried out. The system gives 99.4% overall accuracy and computation time of 0.15 seconds using active contour-based method in combination with NICK's thresholding which makes it a fast, robust and effective SCA detection system.

Different machine learning methods are favourably used by many researchers for accurate SCA detection. Among them widely used are: support vector machine (SVM), K-nearest neighbour (KNN), artificial neural network (ANN), self organising map (SOM), Naïve Bayes (NB), random forest (RF), convolution neural network (CNN) etc. This work proposes a hybrid segmentation based normal and sickle cell RBCs detection approach using extreme learning machine (ELM) classifier. The hybrid segmentation method applies active contour-based segmentation and after that adaptive thresholding method. Here four different types of adaptive (local) thresholding techniques are used one by one for segmentation of blood images and compared the results. The extracted geometrical features of RBCs are used for training, testing and validating the ELM classifier. The performance parameters of ELM classifier show good result with fast computational time. This work is proposed with a motive of presenting a robust, fast and automatic classification method for sickle cells anaemia from microscopic blood smear using hybrid segmentation technique.

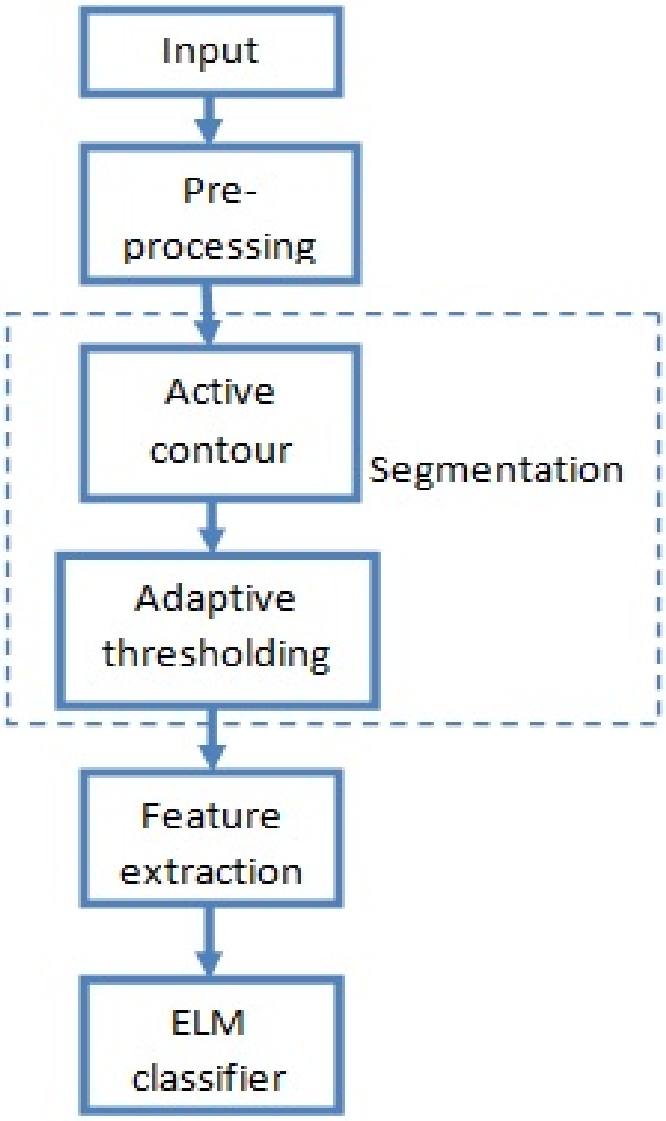


Fig. 6.1 System block diagram for normal RBC and sickle cell classification using ELM

## 6.2 System Block Diagram

The methodology of the proposed technique is illustrated in Fig. 6.1 in the form of a block diagram. The microscopic blood images of the dataset are collected from online library. The database contains ten microscopic images of sickle cell anaemic patient [89]. Both normal

round shaped RBCs and abnormal crescent shaped RBCs are present in the images. First step is to pre-process the images before applying other image processing task.

In this step, artefacts and illumination issue which have been added during image acquisition must be removed. Filtering, debluring, histogram equalization, contrast enhancement etc. are some of the practises that exist for pre-processing of images. After reading the RGB input image in the MATLAB environment, it is transformed to gray scale. Thereafter the image can be filtered using a median filter to remove the impulsive noise present in the image if necessary. Median filter is used here as this filter is useful for preservtion of edges while removing noise.

### 6.2.1 Image segmentation using hybrid method

Segmentation is partitioning an image into its constituent parts which help to examine the image with simpler one. It will split the image into smaller parts (of pixels) and assign labels to each part. The parts with same labels are categorized into a group where they have same properties. Here segmentation process is used to segregate normal and abnormal RBC cells from the background. It is possible to extract features from region of interest after proper segmentation. In this work also after proper segmentation using hybrid method, geometrical or shape features are selected from segmented RBCs. For hybrid segmentation method, active contour-based method and adaptive thresholding are combined and applied on the database.

Some morphological tasks are used after applying the hybrid segmentation method on the raw image to get final form of segmented image. First task is to fill the internal holes present in the cells. Then border of the image is cleaned and incomplete objects are removed. Finally very small undesired objects are removed from the final image. After these tasks, final segmented image is ready and features are extracted from them.

### 6.2.2 Feature extraction from segmented images

Eight geometrical or shape features are extracted here from RBC cells after hybrid segmentation. The extracted features of five normal RBCs and five sickle RBCs are tabulated in Table 6.1. Metric value is used to estimate roundness of cells. For round cells its value is 1 whereas value is in the range $[0.7 - 0]$ for sickle cells. Aspect ratio is also used to form the feature set which is the proportion of length of major axis to that of the minor axis. For normal RBCs, aspect ratio is equal to 1 and for sickle cells the value is much greater than 1. Solidity is defined as the ratio of area of the cell to the area of a convex hull of the cell. Eccentricity is the measure of object's difference from a circular shape [103].

These 8 geometrical features are calculated for each of the segmented image in the database. Using these 8 feature values, feature set is formed for ELM classifier. As four different adaptive thresholding methods are applied here, four different feature set were obtained for each of the image in the database.

### 6.2.3 Classification using ELM classifier

Extreme learning machine (ELM) is emerged as a new technique of machine learning which is proposed by Huang *e*t al. [104]. This feed forward neural network was proposed with single-hidden layer in the paper. They compared classification performance of ELM with other neural network on medical diabetic dataset and found many advantages of ELM over conventional neural networks. ELMs are simple and exceptionally fast where input weights are chosen arbitrarily and analytically determined the output weights. Several researches applied this algorithm for detection of various diseases in medical field with promising results [60, 61, 105]. In [60], authors have described a contour aware segmentation method for WBCs, normal and abnormal RBCs and examined each segmented RBCs. They classified each segmented cells using CNN based extreme machine learning classifier. A comparative

analysis between SVM, KNN and ELM classifiers are carried out by authors in [61] to detect sickle cells. They extracted geometrical and statistical features from segmented cells.

Table 6.1 Shape features of sample normal and sickle RBCs

| Sl. No. | Area | Perimeter | Metric Value | Major Axis | Minor Axis | Aspect Ratio | Eccentricity | Solidity |
|---|---|---|---|---|---|---|---|---|
| Normal cell1 | 419 | 69.94 | 1.07 | 24.96 | 21.45 | 1.16 | 0.51 | 0.97 |
| Normal cell2 | 684 | 90.35 | 1.05 | 30.75 | 28.3 | 1.08 | 0.38 | 0.97 |
| Normal cell3 | 433 | 71.43 | 1.06 | 24.90 | 22.19 | 1.12 | 0.45 | 0.97 |
| Normal cell4 | 458 | 73.69 | 1.05 | 25.32 | 23.11 | 1.09 | 0.40 | 0.96 |
| Normal cell5 | 521 | 78.47 | 1.06 | 26.49 | 25.13 | 1.05 | 0.31 | 0.98 |
| Sickle cell1 | 996 | 127.17 | 0.77 | 52.37 | 25.36 | 2.06 | 0.87 | 0.91 |
| Sickle cell2 | 828 | 121.75 | 0.70 | 49.30 | 23.48 | 2.09 | 0.87 | 0.87 |
| Sickle cell3 | 796 | 125.46 | 0.63 | 46.10 | 23.41 | 1.96 | 0.86 | 0.87 |
| Sickle cell4 | 983 | 144.54 | 0.59 | 56.20 | 25.23 | 2.22 | 0.89 | 0.83 |
| Sickle cell5 | 981 | 144.40 | 0.59 | 61.56 | 22.79 | 2.70 | 0.92 | 0.86 |

Table 6.2 Performance of ELM classifier with hybrid segmentation using Bernsen thresholding

| Sl. No. | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Image1 | 98.21 | 98 | 98 | 97.62 |
| Image2 | 99.48 | 98.48 | 100 | 100 |
| Image3 | 100 | 99.28 | 100 | 98.87 |
| Image4 | 97.80 | 97 | 98.43 | 97.31 |
| Image5 | 98 | 98.20 | 97.17 | 98.37 |
| Image6 | 98.29 | 97.38 | 98 | 98.30 |
| Image7 | 96.42 | 97.78 | 96.48 | 97.41 |
| Image8 | 98.48 | 98 | 97.80 | 98 |
| Image9 | 97.28 | 98.10 | 98.38 | 97.10 |
| Image10 | 98 | 98 | 98 | 97.42 |

## 6.3 Results and Discussion

The microscopic blood smear images of the SCA patient are pre-processed and then hybrid segmentation method is applied to extract the features. Here active contour method is combined with four adaptive thresholding techniques. Table 6.1 shows the range of extracted features for normal and sickle cells of an image considered in the database. The feature vectors generated from the segmented images of the database are employed to train and test the ELM classifier. Whole of the extracted feature sets are divided into 30% and 70% for testing and training purpose respectively. The performance of the ELM classifier is analyzed by evaluating accuracy, sensitivity, specificity and precision of the classifier.

Table 6.2 - Table 6.5 present the performance parameters of ELM classifiers combining active contour-base method with Bernsen, Sauvola, Niblacks and NICK thresholding segmentation method respectively. Table 6.6 demonstrates the overall performance of ELM classifier with hybrid method of segmentation used in the work combining active contour method and

Table 6.3 Performance of ELM classifier with hybrid segmentation using Sauvola thresholding

| Sl. No. | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Image1 | 98.92 | 97.83 | 100 | 100 |
| Image2 | 99.24 | 98.31 | 98.54 | 98.28 |
| Image3 | 97.94 | 98 | 97.24 | 97.85 |
| Image4 | 98.52 | 97.8 | 98 | 97.84 |
| Image5 | 99.24 | 99.37 | 99.25 | 98.72 |
| Image6 | 99.47 | 98.14 | 97.68 | 99.82 |
| Image7 | 98 | 97.28 | 97.10 | 98 |
| Image8 | 99.20 | 98.82 | 98.62 | 97.93 |
| Image9 | 98.42 | 98.36 | 97.79 | 98.18 |
| Image10 | 98.8 | 97.63 | 98.21 | 98.62 |

four different adaptive thresholding techniques for the database. The results show significant performance and better computational time of ELM classifier with hybrid segmentation. The overall performance of ELM classifier is graphically presented in Figure 6.2. Amongst all others, the combination of NICK adaptive thresholding and the ELM classifier gives highest accuracy of 99.4% with lowest computation time of 0.15 second.

The work focuses on an automatic and robust detection process of disorder that is useful for sickle cell anaemic patient. The technique proposed in the work involves hybrid means of segmentation process and ELM classifier for identification of normal and unhealthy or sickle cell RBCs. The hybrid way of segmentation is implemented by means of active contour-based method and four different adaptive thresholding techniques. This proposed methodology yields an efficient sickle cell detection approach for non-overlapping RBCs from microscopic images. Amongst the experimented combinations of the hybrid methodologies, the combination of NICK adaptive thresholding and the ELM classifier offers maximum

Table 6.4 Performance of ELM classifier with hybrid segmentation using Niblack thresholding

| Sl. No. | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Image1 | 98.38 | 97.83 | 97.2 | 98.21 |
| Image2 | 99.68 | 98.38 | 98.6 | 97.73 |
| Image3 | 100 | 99.28 | 100 | 100 |
| Image4 | 98.47 | 98 | 98.03 | 98.49 |
| Image5 | 98.19 | 97.83 | 98.15 | 97.83 |
| Image6 | 99.42 | 99.72 | 97.13 | 97.63 |
| Image7 | 100 | 98.19 | 97.36 | 98.47 |
| Image8 | 98.31 | 99.81 | 98 | 97.41 |
| Image9 | 99.52 | 97.37 | 98.23 | 97.73 |
| Image10 | 97.8 | 98.29 | 97.37 | 97.28 |

accuracy of 99.4% with lowest computation time of 0.15 second which makes it a fast, robust and effective SCA detection system.

Table 6.5 Performance of ELM classifier with hybrid segmentation using NICK thresholding

| Sl. No. | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Image1 | 99.38 | 100 | 99.63 | 98.83 |
| Image2 | 100 | 99.64 | 98.73 | 99.73 |
| Image3 | 98.37 | 98.74 | 97.84 | 98.76 |
| Image4 | 100 | 98.37 | 100 | 100 |
| Image5 | 99.84 | 100 | 99.28 | 100 |
| Image6 | 99.58 | 98.13 | 99.83 | 97.40 |
| Image7 | 98.84 | 98 | 98.94 | 99.29 |
| Image8 | 100 | 99.17 | 100 | 98.94 |
| Image9 | 98.52 | 97.84 | 99.27 | 99.84 |
| Image10 | 99.27 | 98.63 | 98.73 | 97.85 |

Table 6.6 Overall performance of ELM classifier using hybrid segmentation

| Thresholding method | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| Bernsen thresholding | 98.20 | 98 | 98.20 | 98 |
| Sauvola thresholding | 98.80 | 98.20 | 98.20 | 98.60 |
| Niblack thresholding | 99 | 98.50 | 98 | 98.10 |
| NICK thresholding | 99.40 | 98.80 | 99.20 | 99.10 |

Table 6.7 Computation time of ELM classifier with different thresholding method

| Segmentation method | Computation time (sec.) |
|---|---|
| Bernsen thresholding and active contour-based method | 0.49 |
| Sauvola thresholding and active contour-based method | 0.38 |
| Niblack thresholding and active contour-based method | 0.36 |
| NICK thresholding and active contour-based method | 0.15 |

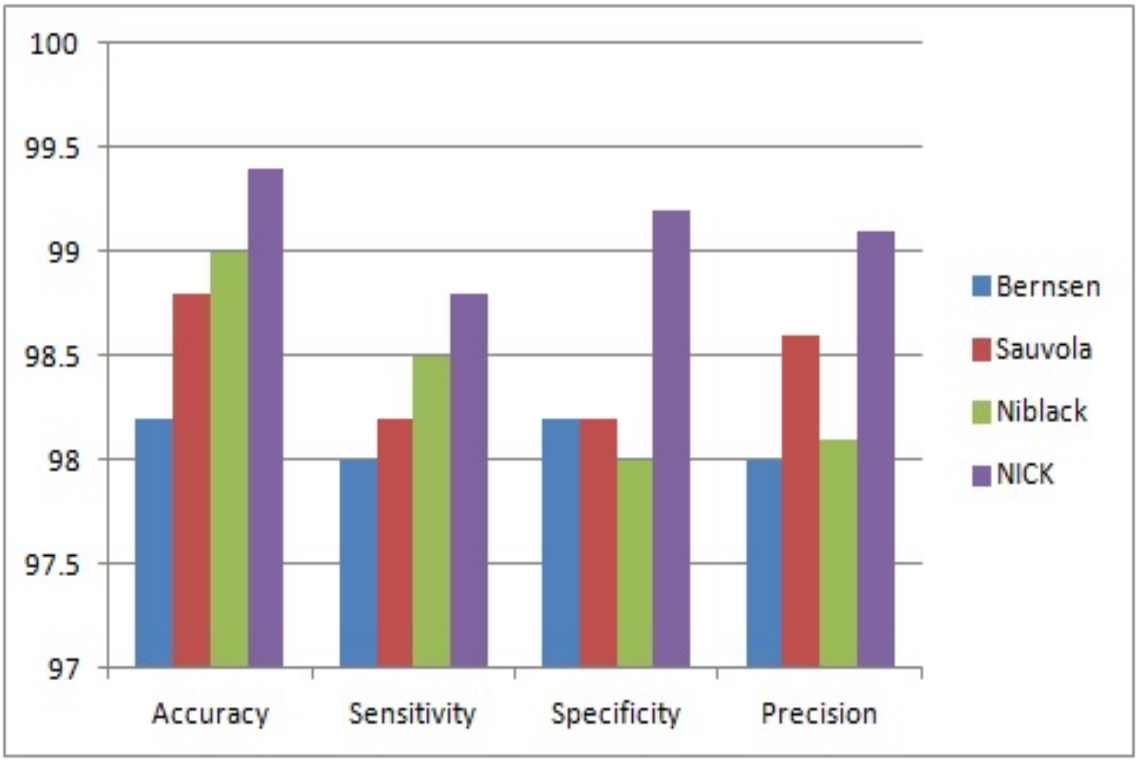


Fig. 6.2 Overall performance of ELM classifier

# 7

# Conclusion

## 7.1 Conclusion

The proposed work focuses on automatic sickle cell detection from microscopic blood images of anaemic person using digital image processing techniques. The use of different image processing techniques replace the time consuming manual counting and detection of abnormal cells by the laboratory technician which give erroneous results upto some extent. The digital image processing methods give accurate results with faster computational time. In this dissertation, we have presented applications of different supervised machine learning

techniques for efficient classification of SCA. These supervised learning algorithms with hybrid segmentation techniques prove to be very effective for classification of SCA.

The summary of our proposed work is illustrated as follows.

- This work mainly proposed application of different adaptive thresholding techniques for SCA detection. The adaptive thresholding methods Bernsen, Sauvola, Niblack and NICK are implemented on image dataset. These adaptive thresholding methods are clubbed with fuzzy C-means clustering and active contour-based technique separately in order to form two varieties of hybrid segmentation methodologies for effective segmentation. After extracting the geometrical features of cells from segmented images, supervised machine learning methods are used for classification of normal and abnormal RBCs.

- Two supervised learning methods, Naïve Bayes and K-nearest neighbor classifier are implemented with first variety of hybrid segmentation methodology comprising adaptive thresholding and fuzzy C-means clustering. This implementation yields a highest accuracy of 98.87% in case of K-nearest neighbor classifier with hybrid combination of segmentation method fuzzy C-means and NICK's adaptive thresholding.

- For improving performance of the system further, second variety of hybrid segmentation methodology comprising adaptive thresholding and active contour-based method implemented with two other supervised learning methods ANN and SVM classifiers. This approach achieves a maximum of 99.2% accuracy in case of SVM classifier with NICK's adaptive thresholding and active contour-based hybrid segmentation method. The performance of ANN classifier is studied for seven different training algorithms by varying the numbers of hidden layer neurons. Out of these seven algorithms, resilient back-propagation algorithm with 10 numbers of hidden neurons gave better performance in ANN with 99% accuracy rate. A comparative analysis is also carried out for ANN and SVM classifier with Otsu's global thresholding method and active contour

method. The performance of ANN and SVM classifiers with adaptive thresholding method is found better than that of global thresholding method.

- The highest accuracy rate of 99.4% is achieved for SCA detection when the network is trained and tested with ELM classifier and hybrid segmentation method comprising NICK's thresholding and active contour is considered.
- A comparison of performance analysis of various methods for SCA detection from microscopic blood images considered in this work is presented in the Table 7.1.

Table 7.1 Comparative performance analysis of different methods considered for the detection of SCA

| Methodology | | Performance parameters (%) | | | |
|---|---|---|---|---|---|
| Segmentation technique | Classification technique | Accuracy | Sensitivity | Specificity | Precision |
| NICK thresholding and Fuzzy C-means clustering | Naïve Bayes | 98.21 | 98.03 | 98.21 | 98.06 |
| | KNN | 98.87 | 98.52 | 98.49 | 98.36 |
| NICK thresholding and Active Contour-based segmentation | ANN | 98.8 | 99.4 | 98.6 | 97.3 |
| | SVM | 99.2 | 98.7 | 98.7 | 98.3 |
| NICK thresholding and Active Contour-based segmentation | ELM | 99.4 | 98.8 | 99.2 | 99.1 |

## 7.2 Future scope

The future directions of the work can be outlined as:

- In pre-processing step, more effective algorithm can be considered for enhancement of images which will enable to produce quality images for further processing.
- In feature extraction step, we have considered geometric or shape parameters of cells to form the feature vector for classifiers. Other types of features can also be considered and experimented with the proposed classification system.
- The application of deep neural network is very much popular in medical image processing field now-a-days. It can be considered for disease detection for large size of dataset. But the numbers of microscopic blood images of sickle cells which are collected from sickle cell anaemic person in online image library is very limited.
- The accuracy of the SCA classifier can be further improved by applying algorithm for proper segmentation of overlapping RBCs present in the microscopic blood images.
- The system proposed in this work is based on detection and classification of abnormal shaped red blood cells from microscopic blood images. This system further can also be implemented for detection of other types of red blood disorder like hemoglobin C disease, hemoglobin S-C disease, various types of thalassemia, spherocytosis etc.

# References


[1] S. Yeruva, M. S. Varalakshmi, B. P. Gowtham, Y. H. Chandana, and P. E. S. N. K. Prasad, Sickle Cell Disease - A Comprehensive Study and Usage of Technology for Diagnosis, International Blood Research & Reviews **11**(2), pp. 6-14, (2020). DOI: 10.9734/ibrr/2020/v11i230125.

[2] Sickle Cell Anemia, Healthline, Available at: https://www.healthline.com/health/sickle-cell-anemia, 2019 (Accessed on 4 August 2019).

[3] D. J. Weatherall, The inherited diseases of hemoglobin are an emerging global health burden, Blood **115**(22), 4331-4336 (2010). DOI: 10.1182/blood-2010-01-251348

[4] Sickle cell disease, WHO Africa, Available at: https://www.afro.who.int/health-topics/sickle-cell-disease, (Accessed on 3 June 2019)

[5] Sickle Cell Disease (SCD), Centers for disease control and prevention, Available at: https://www.cdc.gov/ncbddd/sicklecell/index.html, 2020 (Accessed on 2 August 2019).

[6] E. Wastnedge, D. Waters, S. Patel, K. Morrison, M. Y. Goh, D. Adeloye, I. Rudan, The global burden of sickle cell disease in children under five years of age: a systematic review and meta-analysis, J Glob Health. **8**(2):021103 (2018). DOI: 10.7189/jogh.08.021103.

[7] F. B. Piel, S. I. Hay, S. Gupta, D. J. Weatherall, T. N. Williams, Global burden of sickle cell anaemia in children under five, 2010-2050: modelling based on demo-

graphics, excess mortality, and interventions, PLoS Med. **10**(7):e1001484 (2013). DOI: 10.1371/journal.pmed.1001484

[8] P. Thaker, R. B. Colah, J. Patel, B. Raicha, A. Mistry, V. Mehta, Y. Italia, S. Desai, K. Dave, R. Shanmugam, K. Ghosh, and M. B. Mukherjee, Newborn Screening for Sickle Cell Disease Among Tribal Populations in the States of Gujarat and Madhya Pradesh in India: Evaluation and Outcome Over 6 Years, Front. Med. **8**:731884 (2022). DOI: 10.3389/fmed.2021.731884

[9] A. B. Teli, R. Deori, and S. P. Saikia, Haemoglobinopathies and $\beta$-Thalassaemia among the Tribals Working in the Tea Gardens of Assam, India, Journal of Clinical and Diagnostic Research **10**(12), LC19-LC22, (2016). DOI:10.7860/JCDR/2016/22010.9002

[10] R. B. Colah, M. B. Mukherjee, S. Martin, and K. Ghosh, Sickle cell disease in tribal populations in India, The Indian Journal of Medical Research **141**(5), 509–515, (2015). DOI: 10.4103/0971-5916.159492

[11] V. Raman, T. Seshadri, S. V. Joice, and P. N. Srinivas, Sickle cell disease in India: a scoping review from a health systems perspective to identify an agenda for research and action, BMJ Global Health, **6**:e004322 (2021). DOI:10.1136/bmjgh-2020-004322

[12] G. R. Serjeant, K. Ghosh, J. Patel, Sickle cell disease in India: A perspective, Indian J Med Res. **143** (1):21-4 (2016). DOI: 10.4103/0971-5916.178582.

[13] Sickle cell anemia- Diagnosis & Treatment, Available at: https://www.mayoclinic.org/diseases-conditions/sickle-cell-anemia/diagnosis-treatment/drc-20355882 (Accessed on 3 June 2019)

[14] S. Chakravorty, T. N. Williams, Sickle cell disease: a neglected chronic disease of increasing global health importance, Arch Dis Child. **100**, 48-53 (2015). DOI: 10.1136/archdischild-2013-303773

[15] R. Colah, M. Mukherjee, K. Ghosh, Sickle cell disease in India, Curr Opin Hematol. **21**(3):215-23 (2014). DOI: 10.1097/MOH.0000000000000029.

[16] B. Albayrak, M. B. Darici, F. Kiraci, S. Ogrenci, A. Özmen, and K. Ertez, Sickle Cell Anemia Detection, in *Proc. Medical Technologies National Congress* (2018), pp. 1-4. DOI: 10.1109/TIPTEKNO.2018.8597092

[17] H. A. Aliyu, M. A. Abdul Razak, and R. Sudirman, Segmentation and detection of sickle cell red blood image, AIP Conference Proceedings, **2173**, 020004 (2019). DOI: 10.1063/1.5133919

[18] J. M. Sharif, M. F. Miswan, M. A. Ngadi, M. S. Salam, and M. M. A. Jamil, Red blood cell segmentation using masking and watershed algorithm: A preliminary study, in *IEEE International Conference on Biomedical Engineering (ICoBE)* (2012), pp. 258–262. DOI: 10.1109/ICoBE.2012.6179016

[19] D. K. Das, C. Chakraborty, B. Mitra, A. K. Maiti, and A. K. Ray, Quantitative microscopy approach for shape-based erythrocytes characterization in anaemia, J Microsc **249**(2),136-149 (2013). DOI: 10.1111/jmi.12002.

[20] S. S. Savkare, and S. P. Narote, Blood cell segmentation from microscopic blood image, in *Proc. Int. Conf. Inform. Process.* (2015), pp. 502-505. DOI: 10.1109/INFOP.2015.7489435

[21] N. Ritter, and J. Cooper, Segmentation and border identification of cells in images of peripheral blood smear slides, in *Proc. 13th Australas. Conf. Comput. Sci.* (2007), Vol. 62, pp. 161–169.

[22] H. Berge, D. Taylor, S. Krishnan, and T. S. Douglas, Improved red blood cell counting in thin blood smears, in *Proc. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011), pp. 204-207. DOI: 10.1109/ISBI.2011.5872388.

[23] T. Revathi, and S.Jeevitha, Efficient watershed based red blood cell segmentation from digital images in sickle cell disease, International Journal of Scientific Engineering and Applied Science (IJSEAS) **2**(4), (2016).

[24] M. A. Fadhel, A. J. Humaidi, and S. R. Oleiwi, Image processing-based diagnosis of sickle cell anemia in erythrocytes, in *Proc. IEEE Annual Conf. on New Trends in Information & Communications Technology Applications* (2017), pp. 203-207. DOI: 10.1109/NTICT.2017.7976124

[25] B. Venkatalakshmi, K. Thilagavathi, Automatic Red Blood Cell Counting Using Hough Transform, In *IEEE Conference on Information and Communication Technologies* (2013), pp. 267-271. DOI: 10.1109/CICT.2013.6558103

[26] H. Algailani, E. S. Hamad, Detection of Sickle Cell Disease Based on an Improved Watershed Segmentation, in *Proc. International Conference on Computer, Control, Electrical, and Electronics Engineering* (2018), pp. 1-4, DOI: 10.1109/ICCCEEE.2018.8515855

[27] S. H. Shirazi, A. I. Umar, N. U. Haq, S. Naz, M. I. Razzak, Accurate Microscopic Red Blood Cell Image Enhancement and Segmentation, In: F. Ortuño, I. Rojas (eds) Bioinformatics and Biomedical Engineering. IWBBIO 2015. Lecture Notes in Computer Science, vol 9043. Springer, Cham. DOI: 10.1007/978-3-319-16483-0_18

[28] M. González-Hidalgo, F. A. Guerrero-Peña, S. Herold-García, A. Jaume-i-Capó, and P. D. Marrero-Fernández, Red blood cell cluster separation from digital images for use in sickle cell disease, IEEE J. Biomed. Health Inform. **19**(4),1514-1525 (2015). DOI: 10.1109/JBHI.2014.2356402

[29] A. S. Mohamad, R. Hamzah, A. S. Mokhtar, and J. Satharb, Sickle Cell Disease Verification via Sobel Edge Algorithms for Image Processing, in *Proc. of the International Conference on Engineering Technologies and Technopreneurship* (2017), pp. 1-4. DOI: 10.1109/ICE2T.2017.8215994

[30] P. Rakshit and K. Bhowmik, Detection of Abnormal Findings in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing, Procedia Technology, **10**, pp. 28-36 (2013). DOI: 10.1016/j.protcy.2013.12.333

[31] A. S. Mohamad, N. S. Abdul Halim, M. N. Nordin, R. Hamzah, and J. Sathar, Automated Detection of Human RBC in Diagnosing Sickle Cell Anemia with Laplacian of Gaussian Filter, in *Proc. IEEE Conference on Systems, Process and Control* (2018), pp. 214-217. DOI: 10.1109/SPC.2018.8704128.

[32] N. Z. N. Rashid, M. Y. Mashor and R. Hassan, Unsupervised color image segmentation of red blood cell for thalassemia disease, in *Proc. 2015 2nd International Conference on Biomedical Engineering (ICoBE)* (2015), pp. 1-6. DOI: 10.1109/ICoBE.2015.7235892.

[33] N. S. Aruna and S. Hariharan, Edge Detection of Sickle Cells in Red Blood Cells, International Journal of Computer Science and Information Technologies, **5**(3)(2014), pp. 4140-4144.

[34] K. Alotaibi, Sickle Blood Cell Detection Based on Image Segmentation (2016), Theses and Dissertations. 1116. Available at: http://openprairie.sdstate.edu/etd/1116

[35] W. Delgado-Font, M. Escobedo-Nicot, M. González-Hidalgo, S. Herold-García, A. Jaume-i-Capó, and A. Mir, Diagnosis support of sickle cell anemia by classifying red blood cell shape in peripheral blood images, Med Biol Eng Comput **58**, pp. 1265–1284 (2020). DOI: 10.1007/s11517-019-02085-9

[36] M. Habibzadeh, A. Krzyzak, and T. Fevens, Application of pattern recognition techniques for the analysis of thin blood smear images, J. Med. Informat. Technol. **18**, 29–40, (2011).

[37] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, and J. Zhang, Robust and Efficient Approach to Diagnose Sickle Cell Anemia in Blood, Advances in Intelligent Systems and Computing, Vol. 940, pp. 560-570, Springer, (2020). DOI: 10.1007/978-3-030-16657-1_52

[38] O. D. Fenwa, F. A. Ajala, and O. T. Adedeji, Performance evaluation of selected noise removal algorithms in sickle cell images, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), **4**(1), pp. 1-5 (2015).

[39] M. Veluchamy, K. Perumal, and T. Ponuchamy, Feature extraction and classification of blood cells using artificial neural network, American Journal of Applied Sciences, **9** (5), pp. 615–619, (2012). DOI: 10.3844/AJASSP.2012.615.619

[40] H. A. Elsalamony, Detection of some anaemia types in human blood smears using neural networks, Meas. Sci. Technol. **27**, 085401 (2016). DOI: 10.1088/0957-0233/27/8/085401

[41] M. Khalaf, A. J. Hussain, R. Keight, D. Al-Jumeily, R. Keenan, C. Chalmers, P. Fergus, W. Salih, D. H. Abd, and I. O. Idowu, Recurrent neural network architectures for analysing biomedical data sets, in *Proc.* $10^{th}$ *Int. Conf. Develop. eSystems Eng.* (2017), pp. 232-237. DOI: 10.1109/DeSE.2017.12

[42] H. A. Elsalamony, Sickle Anemia and Distorted Blood Cells Detection Using Hough Transform Based on Neural Network and Decision Tree, in *Proc. Int'l Conf. IP, Comp. Vision, and Pattern Recognition* (2014), pp. 45-51.

[43] R. Tomari, W. N. W. Zakaria, M. M. A. Jamil, F. M. Nor, and N. F. N. Fuad, Computer aided system for red blood cell classification in blood smear image, Procedia Comput. Sci. **42**, pp. 206-213, (2014) DOI: 10.1016/j.procs.2014.11.053

[44] H. A. Elsalamony, Healthy and unhealthy red blood cell detection in human blood smears using neural networks, Micron **83**, 32-41 (2016). DOI: 10.1016/j.micron.2016.01.008

[45] M. Khalaf, A. J. Hussain, R. Keight, D. AI-Jumeily, R. Keenan, P. Fergus, and I. O. Idowu, The utilisation of composite machine learning models for the classification of medical datasets for sickle cell disease, in *Proc. Sixth Int. Conf. Digit. Inform. Process. Commun.* (2016), pp. 37-41. DOI: 10.1109/ICDIPC.2016.7470788

[46] G. Liang, H. Hong, W. Xie, and L. Zheng, Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification, in *IEEE Access*, vol. 6, pp. 36188-36197, (2018). DOI: 10.1109/ACCESS.2018.2846685

[47] Z. Omogbadegun, I. Ogundele, and O. Oladipupo, Applying neural network-based approach to sickle cell disease-related pain classification, Journal of Computer Science **15**(6), 861-872, (2019). DOI: 10.3844/jcssp.2019.861.872

[48] M. Khalaf, A. J. Hussain, D. Al-Jumeily, R. Keight, R. Keenan, P. Fergus, H. Al-Askar, A. Shaw, and I. O. Idowu, Training Neural Networks as Experimental Models: Classifying Biomedical Datasets for Sickle Cell Disease, International Conference on Intelligent Computing, (2016), pp. 784-795. DOI: 10.1007/978-3-319-42291-6_78

[49] J. A. Akrimi, A. Suliman, L. E. George, and A. R. Ahmad, Classification red blood cells using support vector machine, in *Proc. of the* $6^{th}$ *International Conference on Information Technology and Multimedia* (2014), pp. 265-269. DOI: 10.1109/ICIMU.2014.7066642

[50] T. S. Chy and M. A. Rahaman, Automatic Sickle Cell Anemia Detection Using Image Processing Technique, in *Proc. International Conference on Advancement in Electrical and Electronic Engineering* (2018), pp. 98-101, DOI: 10.1109/ICAEEE.2018.8642984.

[51] V. Acharya, and P. Kumar, Identification and red blood cell classification using computer aided system to diagnose blood disorders, in *Proc. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2017), pp. 2098-2104. DOI: 10.1109/ICACCI.2017.8126155.

[52] H. A. Elsalamony, Detection of anaemia disease in human red blood cells using cell signature, neural networks and SVM, Multimedia Tools and Applications **77**(12), pp. 15047-15074, (2018). DOI: 10.1007/s11042-017-5088-9

[53] V. Sharma, A. Rathore, and G. Vyas, Detection of sickle cell anemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing, in *Proc. Int. Conf. Inventive Comput. Technol.* (2016) pp. 1-5. DOI: 10.1109/INVENTIVE.2016.7830136

[54] X. Gual-Arnau, S. Herold-García, and A. Simó, Erythrocyte shape classification using integral-geometry-based methods, Med. Biol. Eng. Comput. **53**(7), pp. 623-633,(2015). DOI: 10.1007/s11517-015-1267-x

[55] V. Acharya, and K. Prakasha, Computer aided technique to separate the red blood cells, categorize them and diagnose sickle cell anemia, Journal of Engineering Science and Technology Review **12**(2) pp. 67-80 (2019). DOI: 10.25103/jestr.122.10

[56] T. H. Lavanya, and S. Sushritha, Detection of Sickle Cell Anemia and Thalassemia using Image Processing Techniques, International Journal of Innovative Science and Research Technology, **2**(12), pp. 399-405, (2017).

[57] M. Xu, D. P. Papageorgiou, S. Z. Abidi, M. Dao, H. Zhao, and G. E. Karniadakis, A deep convolutional neural network for classification of red blood cells in sickle cell anemia, PLoS Comput. Biol. **13** (10)(2017). DOI: 10.1371/journal.pcbi.1005746

[58] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, Deep learning models for classification of red blood cells in microscopy images to aid in Sickle cell anemia diagnosis, Electronics **9**(3), pp. 427 (2020). DOI: 10.3390/electronics9030427

[59] H. A. Aliyu, M. A. Abdul Razak, R. Sudirman, and N. Ramli, A deep learning AlexNet model for classification of red blood cells in sickle cell anemia, IAES International Journal of Artificial Intelligence **9**(2), 221-228, (2020). DOI: 10.11591/ijai.v9.i2.pp221-228

[60] M. I. Razzak, S. Naz, Microscopic blood smear segmentation and classification using deep contour aware CNN and extreme machine learning, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017), pp. 801-807. DOI: 10.1109/CVPRW.2017.111

[61] T. S. Chy, M. A. Rahaman, A comparative analysis by KNN, SVM & ELM classification to detect sickle cell anemia, In *Proc. Int. Conf. Robot., Elect. Signal Process. Techn.* (2019), pp. 455–459. DOI: 10.1109/ICREST.2019.8644410

[62] P. K. Das, S. Meher, R. Panda, A. Abraham, A Review of Automated Methods for the Detection of Sickle Cell Disease, IEEE Reviews in Biomedical Engineering, **13**:309-324, (2020). DOI: 10.1109/RBME.2019.2917780

[63] K. Ghosh, R. B. Colah, M. B. Mukherjee, Haemoglobinopathies in tribal populations of India, Indian J Med Res. **141**(5):505-8 (2015). DOI: 10.4103/0971-5916.159488.

[64] R. Panyang, A. B. Teli, S. P. Saikia, Prevalence of anemia among the women of childbearing age belonging to the tea garden community of Assam, India: A community-based study, Journal of Family Medicine and Primary Care, **7**(4):734-738 (2018). DOI: 10.4103/jfmpc.jfmpc_274_17.

[65] Sickle Cell Anemia Control Program, National Health Mission, Gujarat, Available at: https://nhm.gujarat.gov.in/sickle-cell.htm (Accessed on 13 June 2020)

[66] C. Hockham, S. Bhatt, R. Colah, M. B. Mukherjee, B. S. Penman, S. Gupta, and F.B Piel, The spatial epidemiology of sickle-cell anaemia in India, Scientific Reports **8**, 17685 (2018). DOI: 10.1038/s41598-018-36077-w

[67] Point mutation, Available at: https://en.wikipedia.org/wiki/Point_mutation, 2020 (Accessed on 16 January 2020).

[68] A case study of the effects of mutation: Sickle cell anemia, Available at: https://evolution.berkeley.edu/evolibrary/article/0_0_0/mutations_06, 2020 (Accessed on 2 January 2020).

[69] Sickle Cell Disease, Available at: https://ada.com/conditions/sickle-cell-disease/, 2020 (Accessed on 21 February 2020).

[70] L. K. Lee, S. C. Liew, W. J. Thong, A review of image segmentation methodologies in medical image, Advanced Computer and Communication Engineering Technology. Lecture Notes in Electrical Engineering, Vol. 315, (Springer) (2015). DOI: 10.1007/978-3-319-07674-4_99

[71] W. Khan, Image Segmentation Techniques: A Survey, Journal of Image and Graphics **1**(4), pp. 166-170, (2014). DOI: 10.12720/joig.1.4.166-170

[72] M. Sezgin, and B. Sankur, Survey over image thresholding techniquesand quantitative performance evaluation, Journal of Electronic Imaging, **13**, 146 (2004).

[73] S. Kaushik, An Introduction to Clustering and different methods of clustering, Available at: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/ (Accessed on 3 March 2021).

[74] Fuzzy C-Means Clustering Algorithm, Available at: https://www.datanovia.com/en/lessons/fuzzy-clustering-essentials/fuzzy-c-means-clustering-algorithm/ (Accessed on 8 March 2021).

[75] C. Patgiri, and A. Ganguly, Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve Bayes Classifier and K-nearest neighbor classifier, Biomedical Signal Processing and Control, **68** (2021), 102745. (Elsevier) DOI: 10.1016/j.bspc.2021.102745

[76] M. Kass, A. Witkin, and D. Terzopoulos, Snakes, active contour model, International Journal of Computer Vision, **1** (1988), pp. 321–331.

[77] R. J. Hemalatha, T. R. Thamizhvani, A. Josephin Arockia Dhivya A, Active Contour Based Segmentation Techniques for Medical Image Analysis, Medical and Biological Image Analysis, Robert Koprowski, IntechOpen (2018).

[78] C. Patgiri, and A. Ganguly, Machine Learning Techniques for Automatic Detection of Sickle Cell Anaemia using Adaptive Thresholding and Contour-based Segmentation Method, Asian Pacific Journal of Health Sciences, **9(4)** (2022), pp. 165-170. DOI: 10.21276/apjhs.2022.9.4.33

[79] A. Gautam, P. Singh, B. Raman, and H. Bhadauria, Automatic classification of leukocytes using morphological features and Naïve Bayes classifier, in *Proc. IEEE Region 10 Conference (TENCON)* (2016), pp. 1023-1027.

[80] D. Berrar, Bayes' Theorem and Naive Bayes Classifier, Encyclopedia of Bioinformatics and Computational Biology, **1** (2018), pp. 403-412.

[81] S. Agatonovic-Kustrin, R Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, Journal of Pharmaceutical and Biomedical Analysis, **22**(5), 717-727, (2000).

[82] H. Kukreja, N. Bharath, C. S. Siddesh, S. Kuldeep, An introduction to artificial neural network, International Journal Of Advance Research And Innovative Ideas In Education, **1**(5), 27-30, (2016).

[83] I. Basheer, M. N. Hajmeer, Artificial Neural Networks: Fundamentals, Computing, Design, and Application, Journal of microbiological methods, **43**, 3-31, (2001).

[84] N. H. Alharbi, R. O. Bameer, S. S. Geddan, H. M. Alharbi, Recent Advances and Machine Learning Techniques on Sickle Cell Disease, Future Computing and Informatics Journal, **5**(1), 46-59, (2020).

[85] R. Gandhi, Support Vector Machine — Introduction to Machine Learning Algorithms, Medium, (2018). Available at https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[86] G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: Theory and applications, Neurocomputing, **70**(1–3): 489-501, (2006). DOI: 10.1016/j.neucom.2005.12.126

[87] S. Ding, X. Xu, R. Nie, Extreme learning machine and its applications, Neural Comput & Applic **25**: 549–556, (2014). DOI: 10.1007/s00521-013-1522-8

[88] C. Patgiri, and A. Ganguly, Red Blood Cell and Sickle Cell Detection from Microscopic Blood Images of Sickle Cell Anemic Patient, 2019 IEEE International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)(2019), pp. 474-478. DOI: 10.1109/WiSPNET45539.2019.9032790.

[89] Centers for Disease Control and Preventation Public Health Image Library (online). Available at https://phil.cdc.gov/phil/home.asp.

[90] C. Patgiri, and A. Ganguly, Comparative Study on Different Local Adaptive Thresholding Techniques for Detection of Sickle Cell Anaemia from Microscopic Blood Images, 2019 IEEE $16^{th}$ India Council International Conference (INDICON)(2019), pp. 1-4. DOI: 10.1109/INDICON47234.2019.9030304.

[91] W. Niblack, *An Introduction to Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, (1986).

[92] N. Senthilkumaran, S. Vaithegi, Image Segmentation by using thresholding Techniques for medical images, International Journal of Computer Science & Engineering , **6**(1), (2016).

[93] J. Bernsen, Dynamic Thresholding of Grey-level Images, in *Proceedings of the 8th International Conference on Pattern Recognition* (1986), pp. 1251-1255.

[94] J. Sauvola, and M. Pietikainen, Adaptive Document Image Binarization, Pattern Recognition, **33**(2)(2000), pp. 225-236.

[95] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, Comparison of Niblack inspired Binarization methods for ancient Documents, in *16th International Conference on Document Recognition and Retrieval* (2009).

[96] N. Sharma and L. M. Aggarwal, Automated medical image segmentation techniques, Journal of Medical Physics, 35(1)(2010), pp. 3-14.

[97] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Sys. Man. Cyber, **9**:62-66, (1979).

[98] Chan TF, Vese LA. Active contours without edges. IEEE Trans. Image Process. 2001;10(2):266-277.

[99] Chassery JM, Garbay C. An Iterative Segmentation Method Based on a Contextual Color and Shape Criterion. IEEE Trans. Pattern Anal. Mach. Intell. 1984;6(6):794-800.

[100] Shi Z, He L, Suzuki K, et al. Survey on Neural Networks Used for Medical Image Processing. Int J Comput Sci. 2009;3(1):86-100.

[101] Jiang J, Trundle P, Ren J. Medical image analysis with artificial neural networks. Comput Med Imaging Graph. 2010;34(8):617-631.

[102] N. B. Rais, M. S. Hanif, I. A. Taj, Adaptive thresholding technique for document image analysis, in *Proc. 8th International Multitopic Conference* (2004), pp. 61-66.

[103] D. A. Khusna, H. A. Nugroho, I. Soesanti, Analysis of shape features for lesion classification in breast ultrasound images, AIP Conference Proceedings (2016) **1755**: 160009. DOI: 10.1063/1.4958602

[104] G. B. Huang, Q. Y. Zhu, and C. K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, IEEE International Joint Conference on Neural Networks (2004), pp. 985-990. DOI: 10.1109/IJCNN.2004.1380068

[105] S. H. Shirazi, A. I. Umar, N. Haq, S. Naz, M. I. Razzak, and A. Zaib, Extreme learning machine based microscopic red blood cells classification, Cluster Computing, **21**: 1-11, (2018). DOI: 10.1007/s10586-017-0978-1

# A

# List of Publications

1. **Chayashree Patgiri** and Amrita Ganguly, Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve Bayes Classifier and K-nearest neighbor classifier, Biomedical Signal Processing and Control, **68** (2021), 102745. (Elsevier) DOI: 10.1016/j.bspc.2021.102745

2. **Chayashree Patgiri** and Amrita Ganguly, Machine Learning Techniques for Automatic Detection of Sickle Cell Anaemia using Adaptive Thresholding and Contour-based Segmentation Method, Asian Pacific Journal of Health Sciences, **9(4)** (2022), pp. 165-170. DOI: 10.21276/apjhs.2022.9.4.33

3. **Chayashree Patgiri** and Amrita Ganguly, Comparative Study on Different Local Adaptive Thresholding Techniques for Detection of Sickle Cell Anaemia from Microscopic Blood Images, 2019 IEEE $16^{th}$ India Council International Conference (INDICON), 2019, pp. 1-4. DOI: 10.1109/INDICON47234.2019.9030304.

4. **Chayashree Patgiri** and Amrita Ganguly, Red Blood Cell and Sickle Cell Detection from Microscopic Blood Images of Sickle Cell Anemic Patient, 2019 IEEE International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), 2019, pp. 474-478. DOI: 10.1109/WiSPNET45539.2019.9032790.