# Automated Leukemia Detection using Microscopic Images

*Thesis submitted in partial fulfillment of the requirements for the award of degree of*

**Master of Engineering**

in

**Computer Science Engineering**


*Submitted By*

**Nimesh Patel**

**Roll No. 801332019**

Under the supervision of:

**Dr. Ashutosh Mishra**

Assistant Professor

CSE Department



COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004


**July 2015**

## Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*Automated Leukemia Detection using Microscopic Images*", in partial fulfillment of the requirements for the award of degree of Master of Technology in *Computer Science Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Ashutosh Mishra* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

N. P. Patel

**(Nimesh Patel)**

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

**(Dr. Ashutosh Mishra)**

Computer Science and Engineering Department

**Countersigned by**

**(Dr. Deepak Garg)**

Head

Computer Science and Engineering Department

Thapar University

Patiala

**(Dr. S. S. Bhatia)**

Dean (Academic Affairs)

Thapar University

Patiala

## Acknowledgement

First of all, I am grateful to the God for the good health and wellbeing that were necessary to complete this thesis.

It gives me great pleasure to express my gratitude towards the guidance and help I have received from **Dr. Ashutosh Mishra**, in the Department of Computer Science and Engineering. I am extremely thankful and indebted to her for sharing expertise, and sincere and valuable guidance and encouragement extended to me.

I am immensely grateful to **Dr. Deepak Garg**, Head, Computer Science and Engineering Department for his kind help and cooperation. I take this opportunity to express gratitude to all of the Department faculty members for their help and support. I also thank my parents for the unceasing encouragement, support and attention.

I will be failing in my duty if I don't express my gratitude to **Dr. S.S. Bhatia,** Senior Professor and Dean of Academic Affairs, Thapar University, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable.

Last but not least, I would like to thank my family whom I dearly miss and without whose blessings none of this would have been possible. To my parents, I own thanks for their wonderful love and encouragement. I would also like to thank my brother, since he insisted that I should do so. I would also like to thank my close friends for their constant support.

N. P. Patel

**Nimesh Patel**

# Abstract

The microscopic images of the blood cells are observed to find out many diseases. Changes in the blood condition show the development of diseases in an individual. Leukemia can lead to death if it is left untreated. Based on some statistics it is found that the leukemia is the fifth cause of death in men and sixth cause of death in women. Leukemia originates in the bone marrow. Each bone contains a thin material inside it which is also known as a bone marrow. The components of blood are Red Blood Cells (erythrocytes), White Blood Cells (leucocytes), platelets and plasma. Leukemia is detected only by analyzing the white blood cells. So our study is focused only on the white blood cells (WBCs). The cells in the bone marrow start changing and they get infected and become leukemia or infected cells. These leukemia cells are having strange properties than the normal cells. Their growth is abnormal and survival time is more than the normal cells. They interrupt normal cells to carry out their work. After a certain amount of time normal cells die while leukemia cells don't. The old leukemia cells last for a long time and new leukemia cells produce in an abnormal way. The rate at which the leukemia cells progress is different according to the type of leukemia.

In this work, automated approach of leukemia detection is proposed. In a manual method of Leukemia detection, experts check the microscopic images. This is lengthy and time taking process which depends on the person's skill and not having a standard accuracy. The automated Leukemia detection system analyses the microscopic image and overcomes these drawbacks. It extracts the required parts of the images and applies some filtering techniques. K-mean clustering approach is used for white blood cells detection. The proposed system is successfully implemented in MATLAB.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

RBC: Red Blood Cell

WBC: White Blood Cell

ALL: Acute Lymphocytic Leukemia

AML: Acute Myeloid Leukemia

CLL: Chronic Lymphocytic Leukemia

CML: Chronic Myeloid Leukemia

HSV: Hue Saturation Value

RGB: Red Green Blue

GVF: Gradient Vector Flow

PSO: Particle Swarm Optimization

CART: Classification and Regression Trees

EM: Expectation Maximization

KNN: K-Nearest Neighbors

FAB: French American British

SVM: Support Vector Machine

ALL-IDB: Acute Lymphoblastic Leukemia Image Database

# Chapter 1

# Introduction

The microscopic images of the blood cells are observed to find out many diseases. Changes in the blood condition show the development of diseases in an individual. Leukemia can lead to death if it is left untreated. Based on some statistics it is found that the leukemia is the fifth cause of death in men and sixth cause of death in women. Leukemia originates in the bone marrow. Each bone contains a thin material inside it which is also known as a bone marrow which is shown in the fig. 1.1. The components of blood are Red Blood Cells (erythrocytes), White Blood Cells (leucocytes), platelets and plasma which are shown in the Table 1.1. Leukemia is detected only by analyzing the white blood cells. So our study is focused only on the white blood cells (WBCs).

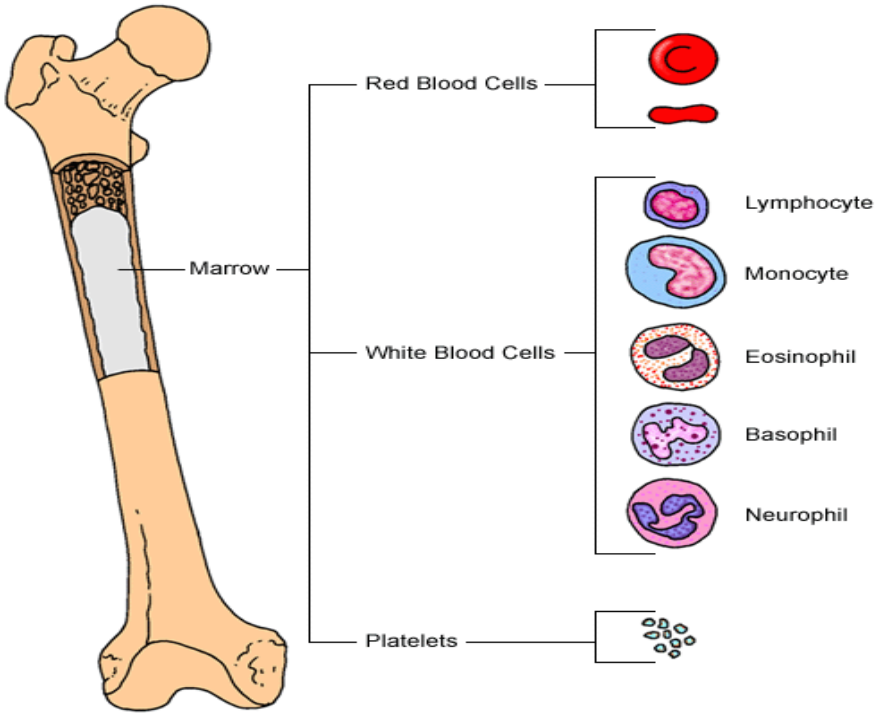



**Figure 1.1:** Bone marrow [1]

The cells in the bone marrow start changing and they get infected and become leukemia or infected cells. These leukemia cells are having strange properties than the normal cells.

Their growth is abnormal and survival time is more than the normal cells. They interrupt normal cells to carry out their work. After a certain amount of time normal cells die while leukemia cells don't. The old leukemia cells last for a long time and new leukemia cells produce in an abnormal way. The rate at which the leukemia cells progress is different according to the type of leukemia.

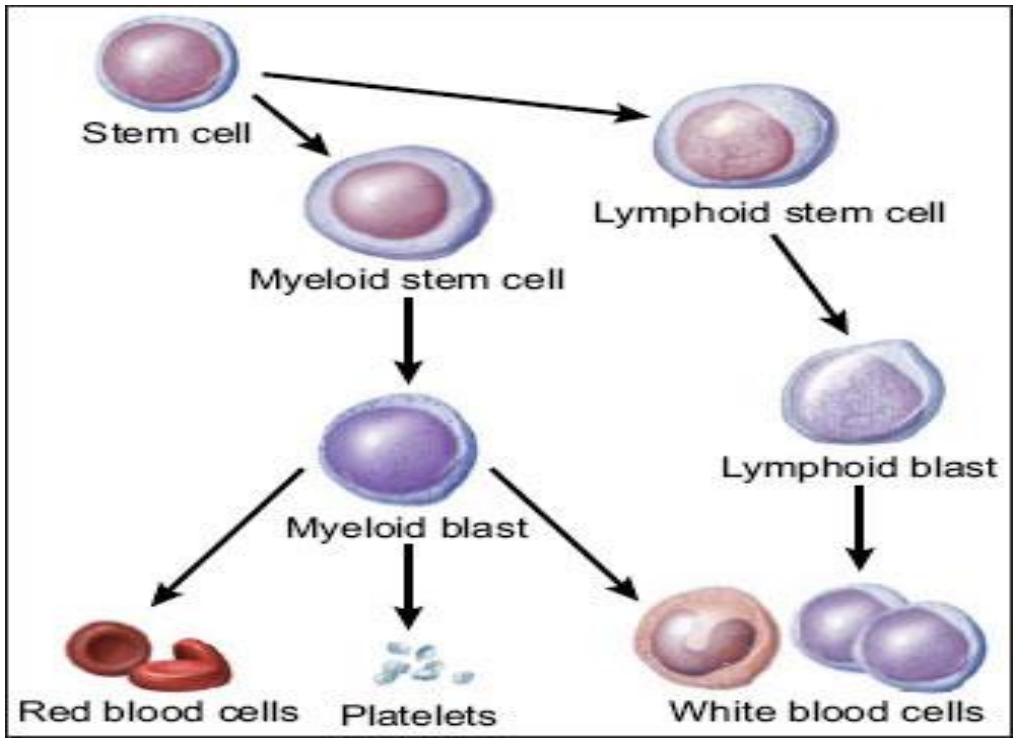
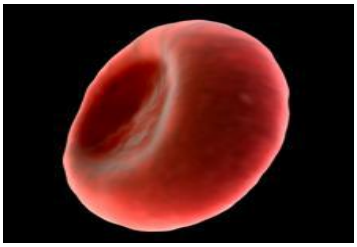
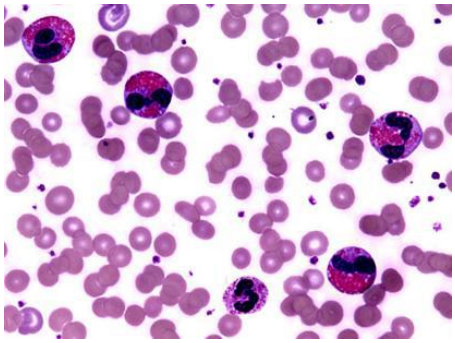
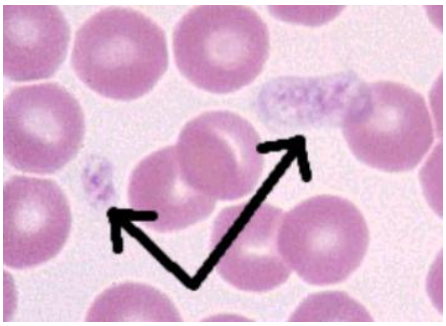
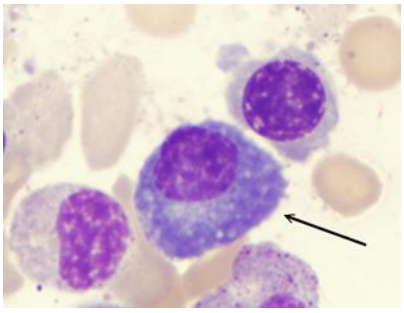



**Figure 1.2:** Stem cell evolution [2]

Fig. 1.2 shows, Stem cell evolves into myeloid stem cell or into lymphoid stem cell. Myeloid stem cell evolves into myeloid blast. Red blood cells (erythrocytes), white blood cells (leucocytes) and platelets are generated from the myeloid blast. Lymphoid stem cell also evolves and leads to the lymphoid blast which will finally generate white blood cells. There exist five types of white blood cells in blood which are lymphocytes, myelocytes, neutrophil, basophil and eosinophil. In leukemia, abnormal white blood cells are being produced by the bone marrow. These abnormal white blood cells should die after some time but they don't and thus they become numerous in count. These numerous abnormal white blood cells interrupt normal white blood cells in doing their work.

**Table 1.1:** Major elements of Blood

| Elements | Description |
|---|---|
| Red Blood Cells (RBCs or erythrocytes)  | Transport oxygen from the lungs to organs and peripheral site. |
| White blood cells (WBCs or leucocytes)  | Defensive role in destroying invading organisms, e.g. bacteria and viruses and assist in the removal of dead or damaged tissue cells. |
| Platelets  | Assist in the clotting process. |
| Plasma  | Carries metabolites antibodies, proteins and nutrients involved in blood clotting. |

## 1.1 White Blood Cells

White blood cells are bigger in size than the red blood cells. The concentration and composition of the white blood cells provide some important information which helps us to find out many diseases. White blood cells can be categorized in to five types: Neutrophil, Basophil, Eosinophil, Lymphocyte and Monocyte which are shown in the Table 1.2. These cells fight against diseases and protect our body.

**Table 1.2(a):** White Blood Cells (Neutrophil, Eosinophil and Basophil)

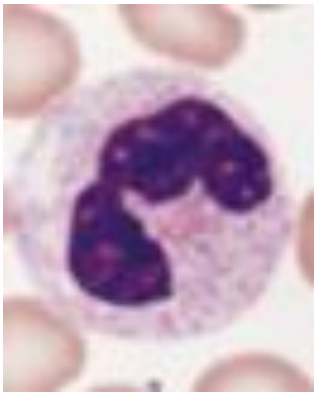
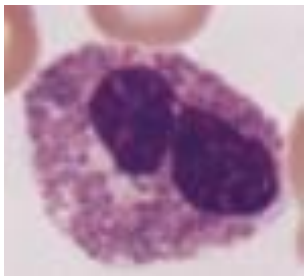
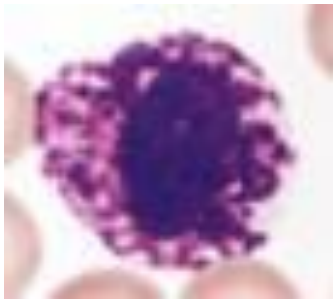
| Types | Description |
|---|---|
| Neutrophil  | This cell is having the nucleus which is containing the cytoplasm. The granules of it are of two types – primary and secondary. Primary granules are seen at the pro-myelocyte stage while the secondary granules are seen at the myelocyte stage. The diameter of it is 12-15 μm. |
| Eosinophil  | These look very similar to the neutrophils. The only change is in the cytoplasmic granules which are red. They insert seditious exudates. They react to the allergies. The diameter of it is 12-15 μm. |
| Basophil  | Basophils can be found only in the normal peripheral blood. Basophils are having more no of cytoplasmic granules in it. These granules overlie nucleus. The diameter of it is 9-10 μm. |

**Table 1.2(b):** White Blood Cells (Monocyte and Lymphocyte)

| Types | Description |
|---|---|
| Monocyte  | Monocytes are generally bigger than the leucocytes. In the bone marrow, the ancestors of the monocytes, promonocytes and monoblasts, are very tough to differentiate from the myeloblasts. Monocytes are present in the bone marrow for very short time. After 20-40 hours they get matured and perform their duties. The diameter of it is 16-20 μm. |
| Lymphocyte  | Lymphocytes are responsible for our body health. They fight against any kind of intruders and infection. This is called are immune system. In the case of any kind of attack, our immune system generates the antigenic specificity to protect our body. The diameter of it is 8-10 μm. |

## 1.2 Types of Leukemia

Leukemia can be classified based upon how fast it becomes severe. Leukemia is classified as chronic or acute.

- Chronic Leukemia – Infected white blood cells perform like normal white blood cells and gradually it increases and becomes severe.
- Acute Leukemia – Infected white blood cells do not perform like normal cells and they increase rapidly in count and become severe.

We can also sub classify it based upon the stem cells generated from the bone marrow.

- Acute Lymphocytic Leukemia (ALL)
- Acute Myeloid Leukemia (AML)
- Chronic Lymphocytic Leukemia (CLL)
- Chronic Myeloid Leukemia (CML)

**Table 1.3(a):** Types of Leukemia (ALL and AML)

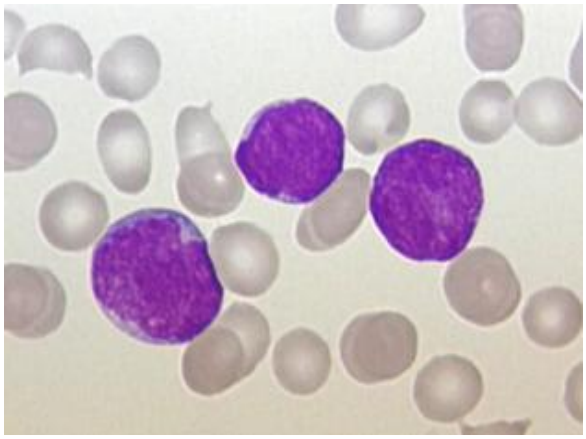
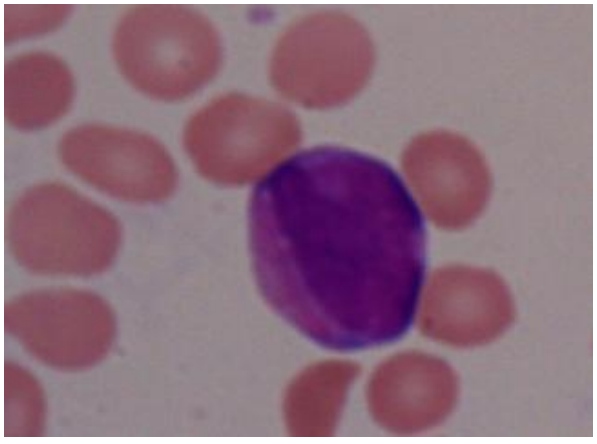
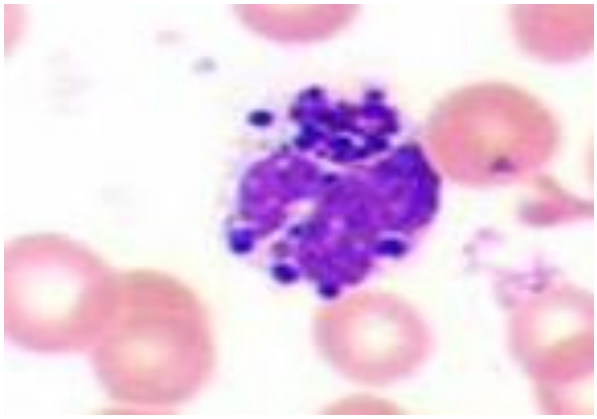
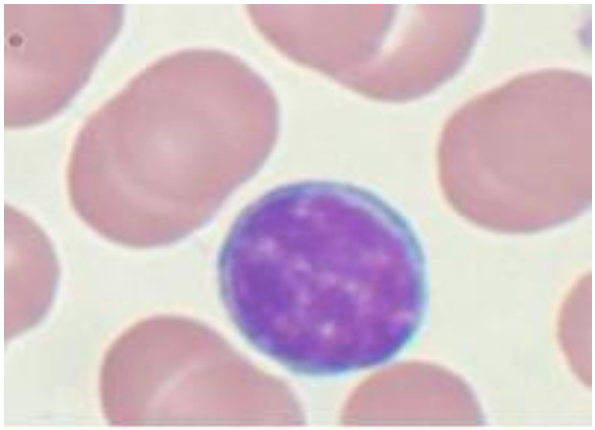
| **Types** | **Description** |
|---|---|
| Acute Lymphocytic Leukemia (ALL)  | It develops in young children mostly. It is also found in adults having age more than 60. |
| Acute Myeloid Leukemia (AML)  | It occurs in children as well as adults. |

**Table 1.3(b):** Types of Leukemia (CLL and CML)

| | |
|---|---|
| Chronic Lymphocytic Leukemia (CLL)<br>  | It mostly occurs in adults. A small number of cases are found in children also. |
| Chronic Myeloid Leukemia (CML)<br>  | It is mostly found in adults more than the age of 55. It is rarely found in young adults. It is never found in children. |

The statistics of leukemia for males and females in UK in 2007 is shown in the table 2.It shows that the rate of survival in 2007 is better than 2001 to 2006. The medical treatments and diagnosis have also improved which is shown in the figure 1.3. The automated detection of leukemia can diagnose the patient early and the necessary treatments can be provided to the patient, so the survival rate can be increased in future.

**Table 1.4:** Leukemia cases in UK for 2007 [3]

| Leukaemia - UK | Males | Females | Persons |
|---|---|---|---|
| Number of new cases (UK 2007) | 4,069 | 2,932 | 7,001 |
| Rate per 100,000 population* | 11.5 | 6.9 | 9.0 |
| Number of deaths (UK 2008) | 2,483 | 1,884 | 4,367 |
| Rate per 100,000 population | 6.4 | 3.6 | 4.9 |
| One-year survival rate (for patients diagnosed 2004-2006, England) | 61% | 62% | - |
| Five-year survival rate (for patients diagnosed 2001-2006, England) | 40% | 41% | - |
| Ten-year predicted survival rate (for patients diagnosed 2007, England & Wales) | - | - | 33.2% |

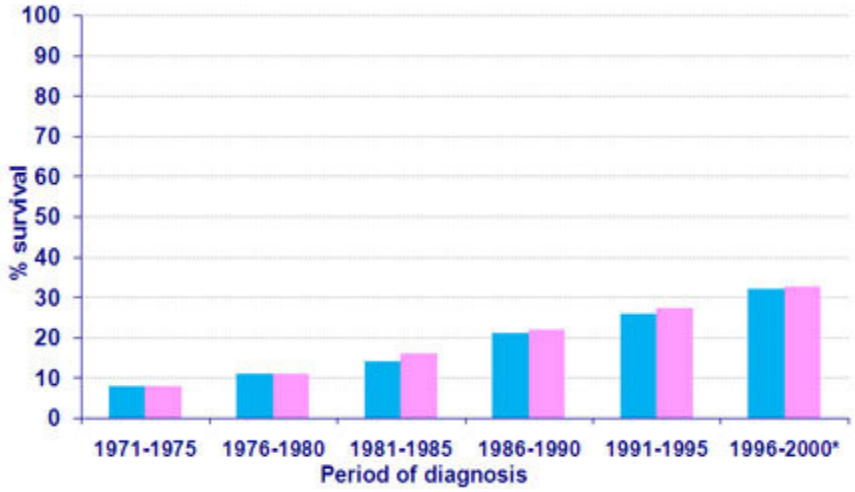



**Figure 1.3:** Leukemia 10-year relative survival rates [3]

### 1.2.1 Acute Lymphocytic Leukemia (ALL)

Acute Lymphocytic Leukemia (ALL) is produced when excessive amount of immature lymphocytes are produced in the bone marrow. These excessive abnormal lymphocytes will interrupt the normal cells to carry out their duties. ALL can lead to death if it is left untreated because the excessive abnormal lymphocyte cells are having some strange properties than the normal cells as these cells have to be died after certain amount of time but they won't unlike normal cells. The features of ALL are completely different than its

myeloid counterpart therefore ALL requires different treatment than AML. According to FAB classification, ALL can be divided into three subtypes: $L_1$, $L_2$ and $L_3$. According to WHO classification, ALL can be divided into three subtypes: pre-B, pre-T and mature-B.

### 1.2.2 Acute Myeloid Leukemia (AML)

Acute Myeloid Leukemia (AML) is produced in the white blood cells. When the non-granular white blood cells develop early and in the abnormal growth, it results in the AML. The blast cells in the bone marrow are developed to form granulocytes which is the white blood cells having small granules or particles. The blast in the stem cells which causes AML doesn't mature. These AML blasts become large in bone marrow and also in the blood. When these cells become large in number, the body cannot stop bleeding and cannot fight against infection. Therefore, the treatment of this disease becomes mandatory as soon as possible after the detection of this disease. To find out the appropriate treatment and the stage of the illness, we have to look at the blast cells in the bone marrow. These abnormal cells are going to be identified using the microscope by clinicians. Depending on the number of blast cells counted and the type of blast in the bone marrow, the disease is classified for the treatment.

The FAB classification was followed to diagnose the AML which is shown in the Table 1 (a) – Table 1 (b), after that the WHO classification came which is shown in Table 2 and now we use the morphological evaluation of the bone marrow and blood.

### 1.2.3 Chronic Lymphocytic Leukemia (CLL)

The Lymphoid blast in the lymphoid stem cells produces B lymphocytes and T lymphocytes. CLL affects the B lymphocytes. These B lymphocytes become abnormal in count in the bone marrow while having CLL.

### 1.2.4 Chronic Myeloid Leukemia (CML)

The granulocytes of almost all types of cells are increased while having CML. The matured Basophils, Myeloid cells and eosinophils are increased specifically in CML.

**Table 1.5(a):** FAB classification of AML (M0 – M3)

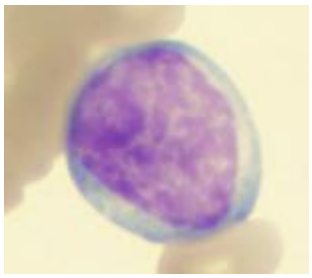
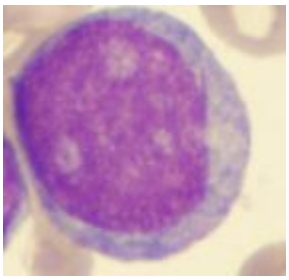
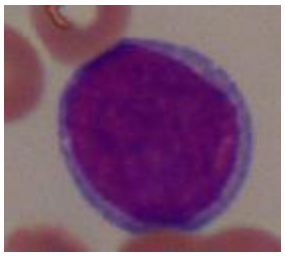
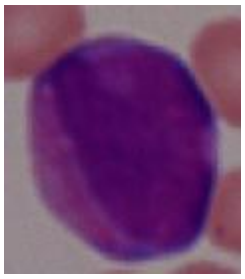
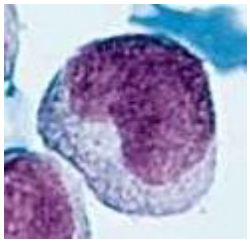
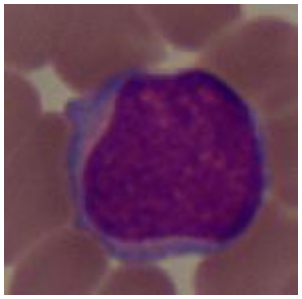
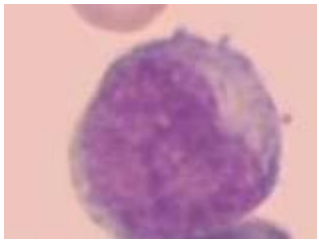
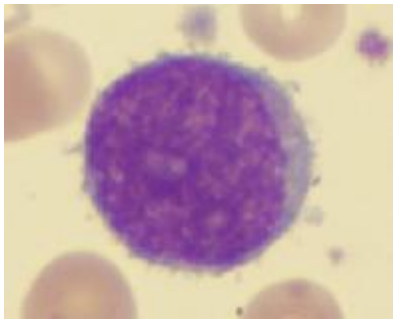
| Category | Morphologic Criteria (Bone Marrow) |
|---|---|
| M0: Minimally differentiated acute myeloid leukemia  | The blast cells are large and quite heterogeneous with the absence of Auer rods |
| M1: Myeloblastic without maturation  | 90% of myeloid cell lines are blasts. The blast cells show few granules but may show Auer rods. |
| M2: Myeloblastic with maturation  | 30-89% of myeloid cell are blasts.<br>>10% are promyelocytes<br>20% are monocytes.<br>Show multiple cytoplasmic granules. |
| M3: Promyelocytic  | Hypergranular promyelocytes with heavy to dust like granules, frequent Auer rods, nucleus often blooded; microgranular variant may occur. Blast cells show multiple Auer rods |

**Table 1.5(b):** FAB classification of AML (M4 – M7)

| Category | Morphologic Criteria (Bone Marrow) |
|---|---|
| M4: Myelomonocytic  | 30-80% of myeloid cell lines are myeloblasts plus maturing neutrophils. 20-80% of myeloid cell lines are monocytic lineage. Blasts have some monocytoid differentiation |
| M5: Monoblastic monocytic  | >80% of a myeloid cell line are monoblasts, promonocytes or monocytes. In M5a 80% of myeloid cell lines are monoblasts; in M5b, <80% were monoblast and the remainder are promonocytes or monocytes |
| M6: Erytheroleukemia  | >=50% of bone marrow cells are erythroid precursors.<br>> 30% of non-erythroid myeloid cell lines are blasts. Showing preponderance of Erythroblast |
| M7: Megakaryocytic  | Blasts in marrow or blood are identified as megakaryocytic lineage. If marrow is undesirable, biopsy shows large tumour of blasts, frequently increased numbers of megakaryocyte and reticulin |

## 1.3 Bone Marrow

Bone marrow can be found inside some large bone in body. It is a special kind of a tissue which contains stem cells. Stem cells are the cells which are transformed into any kind of cells which our body requires. Stem cells are transformed into white blood cells, platelets and red blood cells each of them are having various types of roles to be performed to make the body healthy. Inside the tissue, immature stem cells with extra irons exist. Figure 1.4 shows the cells in the bone marrow while a patient is having leukemia.

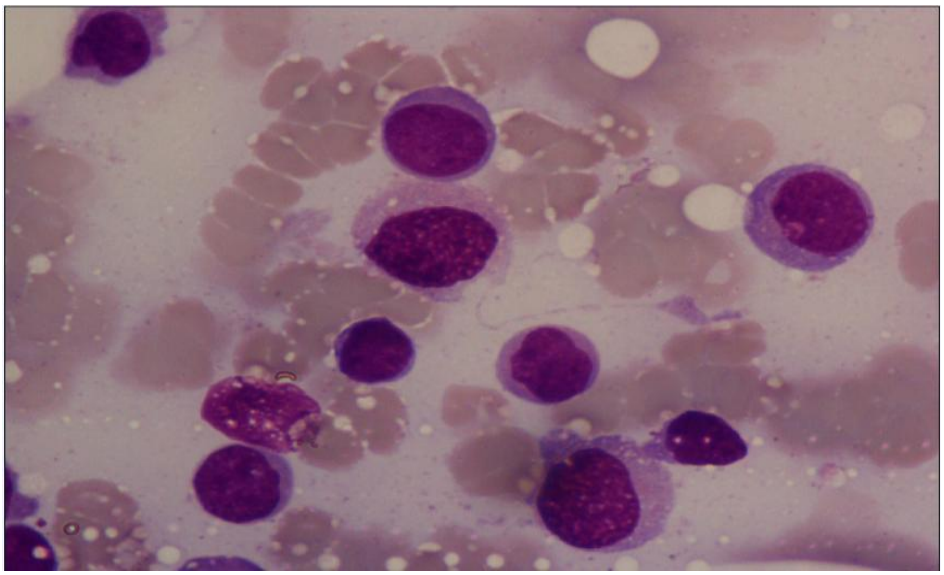


**Figure 1.4:** Bone Marrow sample in M2 AML case [4]

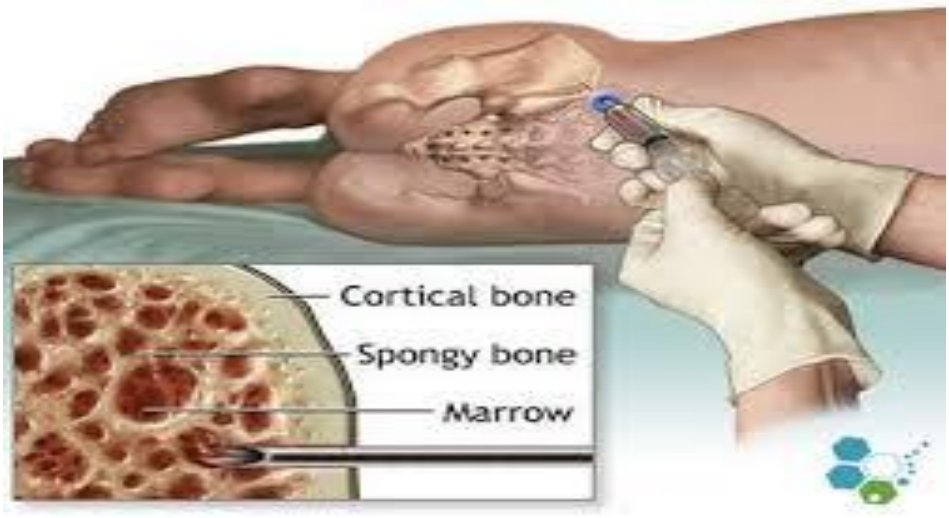



**Figure 1.5:** Blood taken from Bone Marrow [5]

When there are any cells in a body which are weak, abnormal or damaged then these cells need to be replaced with the new cells until this stem cells remain undifferentiated. The new cells need to be replaced to make the body healthy can be white blood cells, platelets or red blood cells. The blast in the stem cell makes myeloid stem cells and lymphoid stem cells. The myeloid blast in the myeloid stem cell makes red blood cells and platelets while the lymphoid blast in the lymphoid stem cell makes white blood cells. Figure 1.5 shows how the bone marrow is taken out from the bone.

# Chapter 2

# Literature Review

In the literature, some has done a valuable work in making the automated system of detecting the leukemia from the microscopic image.

Madhloom [6] performed some image arithmetic operations and some threshold operations to find out the white blood cell nuclei. The challenging task in developing this system was the selection of an appropriate threshold for the segmentation. The threshold used in the system developed is not providing efficient result for the segmentation of white blood cells nuclei.

Kovalev [7] developed a system to classify five types of leucocytes from the blood image. In this process, he detected nuclei first and then applied a region growing techniques to find out the entire membrane. The results achieved were quite good.

Scotti [8] used some threshold operations, low-pass filter for the removal of background and clustering for white blood cells segmentation. Scotti tried to achieve a good segmentation results for the images which are taken under different lightening conditions.

Piuri [9] performed white blood cell segmentation. He used edge detection technique for the each leucocyte. He used a neural network for the classification. He trained neural network by morphological features to recognize lymphoblast.

Halim [10] proposed an automated system which counts number of blasts in the microscopic blood image. He applied some threshold operation on S component of HSV color space to detect white blood cells from the image. The results achieved were quite amazing but the problem in the system is that no method is mentioned for selecting the optimum threshold for the better segmentation. There is no features extracted and no classifier has been used.

Mohapatra [11] applied clustering for white blood cells segmentation and extracted the features like shape, color, texture, fractal, Fourier descriptors and contour. The system

was trained to recognize Leukemia. The results achieved were quite good but the proprietary data set was used to achieve it therefore, it cannot be compared with other methods.

Donida Labati [12] proposed the data set including the blood samples of the normal patients and leukemia patients which found very helpful for our proposed system testing. This data set is publicly available for the research purpose.

## 2.1 Blood cell Research

Some research related to the blood cell has been carried out to identify the blood cells automatically and then after it becomes easy to diagnose the patient. Some of the researchers have developed the system to correctly classify and identify the malaria parasite from the blood cells' microscopic images. The main challenge in developing this system was to segment the microscopic image. They used morphological approach in the developed system.

Liao and Deng [13] developed a system to segment white blood cells from the image. They applied some threshold techniques and then after they applied contour identification. They have assumed that all cells are circular to make the algorithm works efficiently. Due to the assumption of the circular shaped cells, this system is not suitable for the lymphoblast cells which are irregular in shape.

Angulo et al. [14] proposed a system in which he proposed "two-stage blood image segmentation algorithm". They are using binary filtering and some automatic threshold techniques. This system performs well for extracting the nucleus, cytoplasm and nucleolus from the lymphocyte images. The two stage segmentation process has been applied here and because of this the computation time is higher. The images are taken under different lightening condition which makes difficult to choose the optimum threshold for segmentation.

Sinha et al. [15] proposed a scheme which segments the leucocytes automatically. He used EM algorithm and Gaussian mixture modeling. In this method, parameter tuning is not required. This is unsupervised approach. This scheme is not work for all stains.

Umpon [16] invented a technique for the white blood cell nucleus segmentation. He applied fuzzy clustering. This technique works well for the nucleus segmentation but the

cytoplasm extraction is also as important as the nucleus segmentation which is not taken care in this technique.

Dorini *et al.* [17] proposed a scheme for the nucleus extraction. The watershed transform has been used in this scheme which is based on the image forest transform. He has extracted cytoplasm by using the size distribution information. This scheme is not working well if the cytoplasm isn't round.

Dorin Comaniciu et al. [18] proposed an algorithm for the segmentation of cell. The algorithm uses L*U*V color space and divides the image into clusters. This technique works well for the nucleus segmentation but the cytoplasm extraction is also as important as the nucleus segmentation which is not taken care in this technique.

Yang et al. [19] proposed a model to segment leucocytes using GVF. He has used L*U*V color space in the model. The $L_2E$ estimation technique and color gradient has been used in the GVF model. The proposed new GVF model performs well than the standard GVF model but the textures and weak edges are not distinguished properly therefore the lymphocytes cannot be segmented properly.

Yi et al. [20] proposed a neural network trained by PSO to segment the white blood cells from the image. The uniform sampling and mean shift are used to reduce the time and size of the training data set. This scheme doesn't work well for the cytoplasm nucleus detection.

Ghosh et al. [21] introduced a technique to find out accurate threshold for the segmentation of the leucocytes. He used fuzzy divergence in that technique. He has used various functions like Gaussian, Gamma, Cauchy, etc. in that technique. This technique works well for segmenting the nucleus but the extraction of cytoplasm has not been taken care which is also as important as the nucleus extraction in cancer detection.

Shitong [22] developed a technique merging the threshold segmentation, fuzzy and some mathematical morphology. It is very good in detecting the leucocytes faster than any other technique does. The problem with this technique is that it is not separating the nucleus and cytoplasm properly.

Chinwaraphat et al. [23] invented a technique by modifying the fuzzy c-mean clustering technique. The modification has improved the clustering just by reducing the uncertainty between the nucleus and cytoplasm selection. The problem with this improved technique

that we have to do some manual cropping. The improved technique has only been compared to the old technique of Fuzzy c-Means.

Meurie et al. [24] developed a technique which segments cells by using the combined pixel classification. There are so many classifiers used in this technique but the result achieved is not good as it supposed to be. The time of execution is high because of the use of so many classifiers.

Ghosh et al. [25] developed a technique by using the watershed segmentation which extracts the white blood cells from the background. This technique does well for the extraction of the white blood cells from the background but it fails to extract the nucleus and cytoplasm from the background. There is no specific procedure used to find out the threshold for the separation of the nucleus from the cytoplasm.

Ko et al. [26] developed a hybrid technique for the segmentation of the leucocytes. The mean shift clustering and the GVF snake model for the removal of boundary have been used in this technique. For the nucleus and cytoplasm extraction, he has used two different methods independently. The problems with this technique are that the accuracy of segmentation is not good and the execution time is high.

## 2.2 Image Processing Based Methods

Serbouti et al. [27] has used the classification and regression trees (CART) statistical software for the classification. The problem with this scheme is that the identification of the lymphocytes from the lymphoblast in the stem cells as we talked earlier is not proper. Moreover there is no clarification about the features involved and the segmentation scheme.

Foran et al. [28] has invented a scheme to distinguish between the leukemia and the lymphoma. He has achieved an accuracy of 83% in this classification. This scheme has been tested on the 19 cases and found working properly but the problem with this scheme is the cases tested are small in number.

Scotti [29] developed a technique for the classification of Acute Lymphocytic Leukemia. He tested this technique on 150 images and concluded that the lymphoblast can be detected easily with the help of morphological features. He used Otsu thresholding for

segmenting images. For classifying the features, he has used feed forward neural network which may be the reason behind the low recognition power of the system.

Markiewicz et al. [30] researched on the bone marrow images and developed a system which can detect the myeloid blast cells in the bone marrow image shown in the figure 5. The system works quite well for the recognition of the myeloblast up to certain extent but it has never been tested for the identification of the lymphoblast.

Halim et al. [31] developed a method for the counting of blast in the microscopic blood image. He has used HSV color space. The histogram threshold technique is used on the S component of HSV. He applied morphological erosion for segmenting the image. The separation of the nucleus from the cytoplasm is very essential step in the leukemia detection but there is no mentioned method to find out the threshold to do so. Moreover there is no information given about the features and the classifiers used in leukemia identification.

Seshadri et al. [32] developed computer morphometric system in the classification of FAB. A computer based morphometric system has been developed to measure the cell morphology with the help of plotter and digitizer tablet. A digitizer cursor has been used to draw the nucleus, cell and nucleoli in order to calculate the features like perimeter and area. Some of the features like basophilic nature of the cytoplasm and density of the nucleus cannot be measured by this computer based system because computational resources are limited. The main thing about this automated system is that it works well for separating the $L_1$ and $L_2$ samples.

Angulo et al. [33] [34] implemented a watershed transformation for segmenting the lymphocytes. By using some of the cellular typology lymphocytes are classified with the help of morphological features. This method works well although it has never been applied for the classification of lymphoblast nor for segmenting lymphocytes.

Gupta et al. [35] developed a system using relevant vector for identifying lymphoblast. The system identifies three types of lymphoblasts. This system works well for the ALL of children but not for ALL of adults. The problem with this system is that the Otsu's algorithm has been used for segmenting the lymphoblast which is not strong method.

Escalante et al. [36] invented a scheme for classifying the leukemia using the swarm model. The leukemia cells need to be isolated manually to make this system work. These

isolated cells are then segmented by Markov random fields. These segmented nucleus and cytoplasm are then used to find out features of the type of the leukemia i.e. AML vs. ALL, AML subtyping and $L_1$ vs. $L_2$. This scheme works well for the AML vs. ALL classification. The problem with this scheme is that we have to select the region of interest manually.

# Chapter 3

# Problem Statement

It has been observed from the literature that much number of schemes have been developed in counting the blood cells automatically from the microscopic images. Many people are still working in the automated blood cells counting.

The literature on the leucocyte segmentation, it is noticed that large number of methods are only working on the extraction of nucleus but there are very few methods available which are extracting the cytoplasm and even with less accurately. The main reason behind the less accuracy in the cytoplasm extraction is that most of the researchers are using the grey level colour for the extraction of cytoplasm which is not easily separable from the other colours.

It is noticed in the literatures that different approaches are used for the white blood cells detection. Some have used KNN approach, threshold techniques, EM algorithm, Fuzzy rules, watershed transform, GVF model, trained neural network, Fuzzy c-mean clustering, computer morphometric system and many more.

From the literature studied, it has been observed that there are many ways we can make a better system for the identification of leukemia from the microscopic blood image. None of the researchers has used the K-mean clustering for the segmentation of the white blood cells from the microscopic blood image.

In this thesis, K-mean clustering approach has been used on the clean microscopic blood image followed by image cleaning and the extraction of the nucleus and cytoplasm with a good accuracy.

# Chapter 4

# Proposed Work

The proposed system of automated leukemia detection from microscopic image is shown below in fig. 4.1.

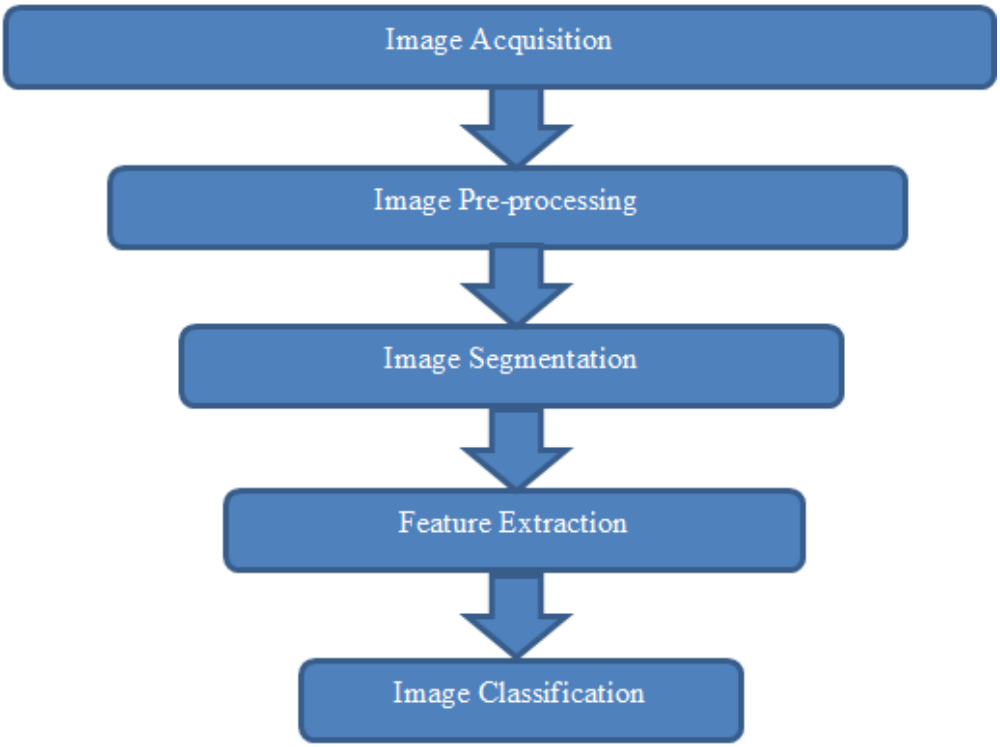



**Figure 4.1:** Proposed system of automated leukemia detection

## 4.1 Image Acquisition

Blood images of the good pixel quality are obtained from any nearest hospital.

## 4.2 Image Pre-processing

The acquired image may contain some noise. There may be a blurred region in the image which is important for our study. The noise is removed from the image using median filtering. Wiener filter is used to remove the blurriness in the image. Image cleaning also need to be performed. In the image cleaning, all the leucocytes which are at the edge of

the image and all the other components which are not leucocytes are to be removed for the better study. Solidity need to be measured for image cleaning. First, area and the convex area of each leucocyte need to be measured and then only we can find solidity for the image cleaning.

$$solidity = \frac{area}{\text{convex area}}$$

The image we have got is in the RGB form which is needed to be converted into the grey scale image for further processing.

## 4.3 Image Segmentation

In this phase our main aim to identify the white blood cells. There exist five types of white blood cells as discussed above but our main study is only on lymphocytes and myelocytes. So, we group only lymphocytes and myelocytes and other three white blood cells like neutrophil, basophil and eosinophil, are discarded from our images. We have applied K-mean clustering for the white blood cells detection. When we convert the images into the grey scales then the nucleus of white blood cells become the darkest region of the image. We have applied some technique for the identification of the grouped leucocytes which is discussed in section 4.3.1. After the identification process, cleaning of the image has been carried out which is discussed in section 4.3.2. Finally, the nucleus and cytoplasm are extracted from the lymphocytes which are discussed in section 4.3.3.

### 4.3.1 Identification of grouped leucocytes

One of the main problems in analyzing the blood image is the adjacent cells. If the cells are grouped or not separated from each other then we cannot study some of the features of cell. In the adjacent cells, their nucleus will be joined therefore we cannot find out area of the nucleus. We have to separate these grouped leucocytes before studying them further. There are so many methods available which are helpful to find out these grouped leucocytes from the blood image. We have used roundness measure to find out these grouped leucocytes. The reason we chose roundness is that we can identify grouped leucocytes just by analyzing the shape of them. Most of the cells will be round in shape

but the grouped cells are not having the round shape. Roundness checks whether the shape is circular or not by excluding the local irregularities. Roundness can be gained by dividing the area of a circle to the area of an object by using the convex perimeter.

$$Roundness = \frac{4 \times \pi \times \text{area}}{convex\_perimeter^2}$$

The value of roundness is 1 if the object is circular and the value of roundness is less than 1 for the non-circular objects. Roundness is not very much sensitive to the irregular boundaries because it excludes the local irregularities. After some well observations we found that the value 0.80 can be used as a threshold to properly distinguish between the single leucocyte and the groups of leucocytes. The components which are having the roundness value more than the value of threshold are considered as the individual leucocyte while the components which are having the roundness value less than the value of threshold are considered as grouped leucocytes. The individual leucocytes are sent next for the further study and the grouped leucocytes can be either sent to the separation process or just can be rejected from our further study. In our model, we reject these grouped leucocytes from our further study. Figure 4.2 shows the identified grouped lymphocytes.

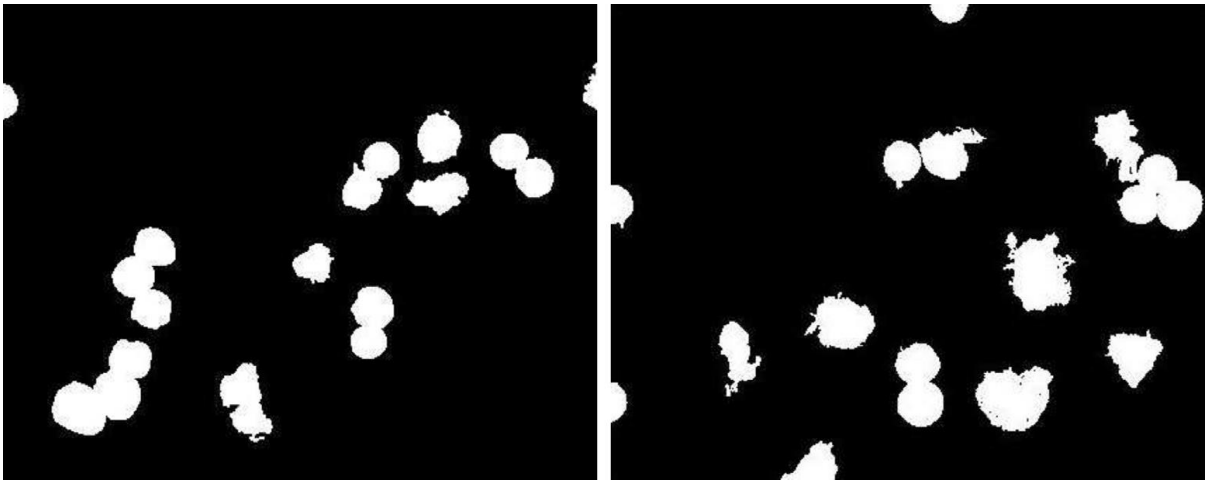


**Figure 4.2:** Leucocytes identified as grouped

### 4.3.2 Image cleaning

The main object which we study in microscopic image is the leucocytes. We need to neglect everything other than leucocytes from the image. When we take a picture of the blood there may be a case that some of the leucocytes are on the edge of the image therefore a portion of the leucocyte appears on the edge of the image. These partial leucocytes may create errors in the study. In the image cleaning process, all the objects which are not leucocytes and the leucocytes which are on the edge of the image are removed so that we can get better results. There are two operations which we need to perform here – "Cleaning the edge of an image" and "Remove abnormal objects". The first one is easier than the second one. We have to count the number of leucocytes. The size of the area is the important measure which needs to be calculated. The mean area is going to be calculated from the size of the area. The mean area is used to discard the irregular components. The component which is having a very small area might the component which is located on the edge of the image. The component which is having very large area might be the adjacent cells of leucocytes. The area and the convex area both need to be calculated for the removal of the components having small and very large area. Solidity is used to find out the density of a component. Solidity value can be obtained by dividing the area to the convex hull of each component.

$$solidity = \frac{area}{\text{convex area}}$$

If the solidity value is 1 then we can say it is a solid object. If the solidity value is less than 1 then we can say it is a component having irregular boundaries. The threshold value for solidity which is used for identifying the abnormal components can be obtained from the image which is having individual leucocytes only. After having so many experiments, value 0.90 can efficiently be used to find out abnormal components from the image. So, 0.90 is the threshold value for the solidity. The components which are having the solidity value less than the threshold are removed. In fact, the components having lesser value than the threshold are the components which are on the edge of the image which need to be discarded. Figure 4.3 shows the cleaned image which can be obtained by removing the leucocytes on the edges.

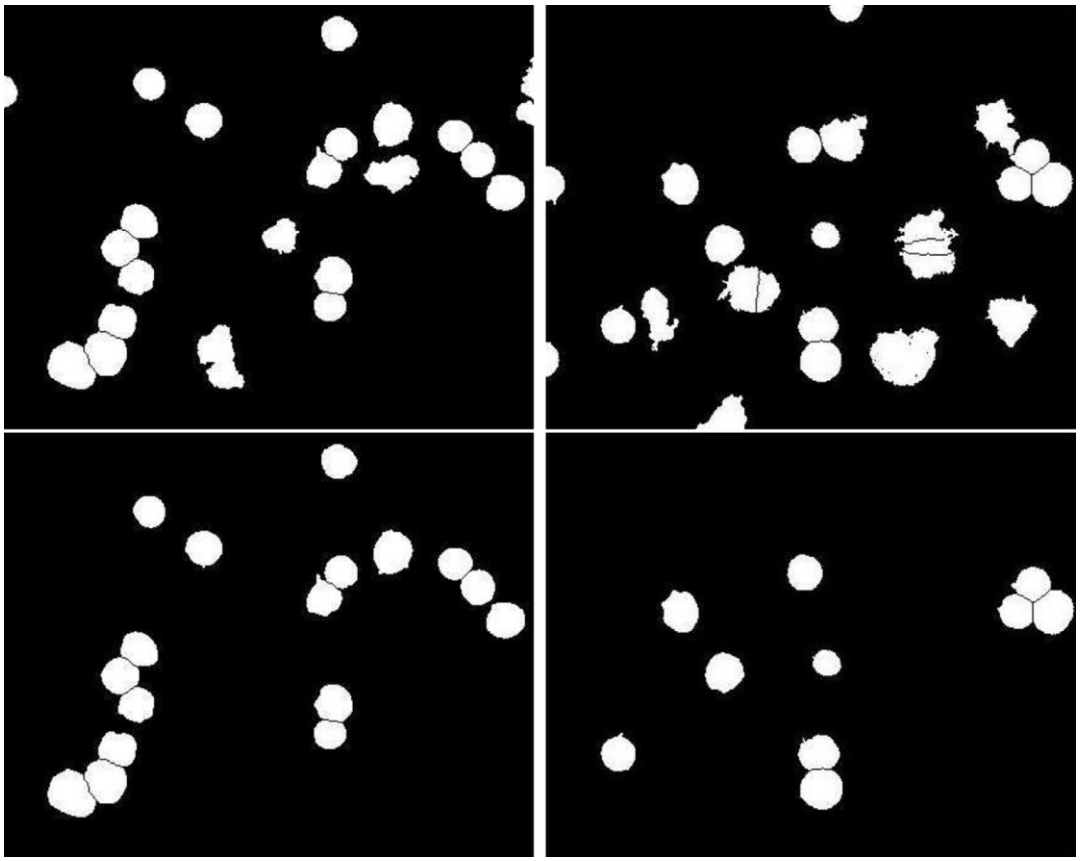


**Figure 4.3:** Final separation results and image cleaning results

### 4.3.3 Nucleus and cytoplasm selection

The leucocytes identified in above steps can now be used to extract the nucleus and cytoplasm. To carry out this step, we crop the image with the bounding box size. This size is the rectangle which can properly fit the component so that we can isolate each components of an image. We have to separate out each leucocyte by this method. The borders of every sub-image obtained like this have to be cleaned up before we proceed. Now the portion outside the leucocyte has to be cropped which will help us in getting the cytoplasm. This method completely removes the artefacts. We have used Cseke's observation to find out nucleus in our method. The observation says that the white blood cells nuclei are more in contrast on the green component of the RGB color space. So, we can get nucleus by using the threshold. To get the cytoplasm we perform subtraction operation between Figure 4.4 (c) and Figure 4.4 (d). The last two figures 4.4 (d) and 4.4 (e) show the nucleus and cytoplasm respectively.

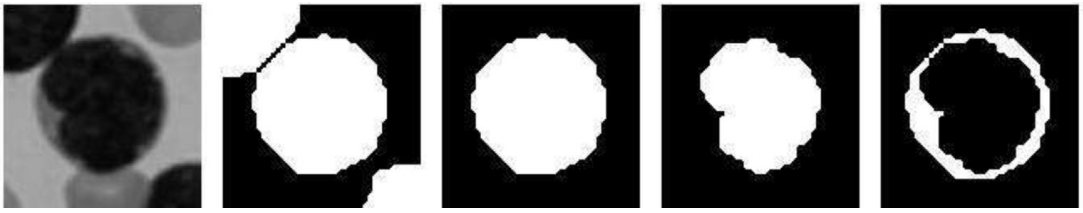


**Figure 4.4:** Left to right: grey level sub-image [a], binary sub-image [b], whole leucocyte sub-image [c], nucleus sub-image [d] and cytoplasm sub-image [e]

## 4.4 Feature Extraction

In this phase we try to extract some of the features from the processed image. Here, we try to find out the features of the nucleus of myelocytes and lymphocytes. Feature extraction is the process of converting the image into data so that we can check these values with the standard values and finally we can differentiate between the cancerous and non-cancerous data. Some of the features which are necessary to be calculated are listed below.

- Color Features – The mean color values of the grey images are acquired.
- Geometric Features – The perimeter, radius, area, rectangularity, compactness, convexity, concavity, symmetry, elongation, eccentricity, solidity are obtained.
- Texture Features – The entropy, energy, homogeneity, correlation are obtained.
- Statistical Features – The skew ness, mean, variance and gradient matrix are obtained.

$$\text{Elongation} = 1 - \frac{minor\ axis}{major\ axis}$$

$$Eccentricity = \frac{\sqrt{(major\ axis^2 - minor\ axis^2)}}{\text{major axis}}$$

$$Rectangularity = \frac{\text{area}}{major\ axis\ \times minor\ axis}$$

$$\text{Convexity} = \frac{perimeter_{convex}}{\text{perimeter}}$$

$$Compactness = \frac{4 \times \pi \times \text{area}}{perimeter^2}$$

Elongation shows the way of an object elongation. Rectangularity shows how well the bounding box is filled. Eccentricity is the ratio of the major axis length and the foci of the ellipse. Convexity shows the relative amount of difference of object from its convex object. Compactness is the ratio of the area of an object and area of circle having same perimeter.

## 4.5 Image Classification

In this final phase, the features extracted are used to provide the final answer. All feature extracted are listed into the different columns with their values. When we give any image as an input to the proposed system then we first calculate the feature values. The values of the test image features are checked with the previously calculated values Based on the values of the input image the SVM classifier classifies that test image into either infected or not infected class.

# Chapter 5

# Experimental Results

The microscopic image has been sent to the proposed system. The system then gives the subsequent images as the result.

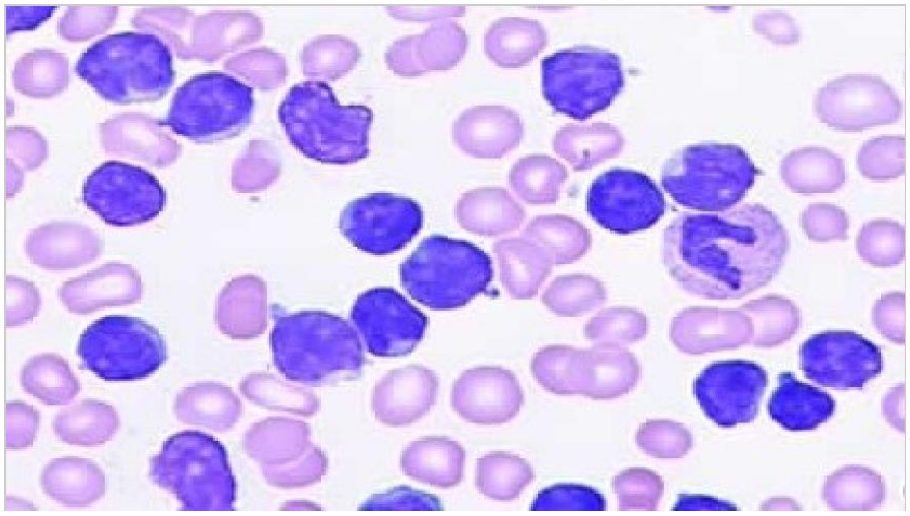


**Figure 5.1:** Original microscopic blood image

Figure 5.1 is the original microscopic image of the blood sample obtained from any nearby hospital. This is given to the proposed system as an input.

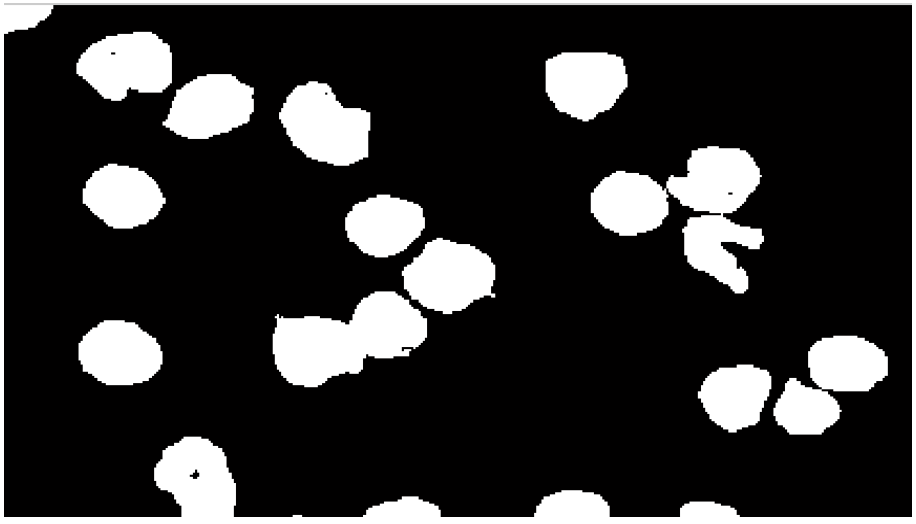


**Figure 5.2:** Grey scale image

The image we got as input is modified by the system by removing the noise. The identification of grouped leucocytes and image cleaning operations are performed on the image and then after it is going to be converted into the grey scale image which is shown in the figure 5.2.

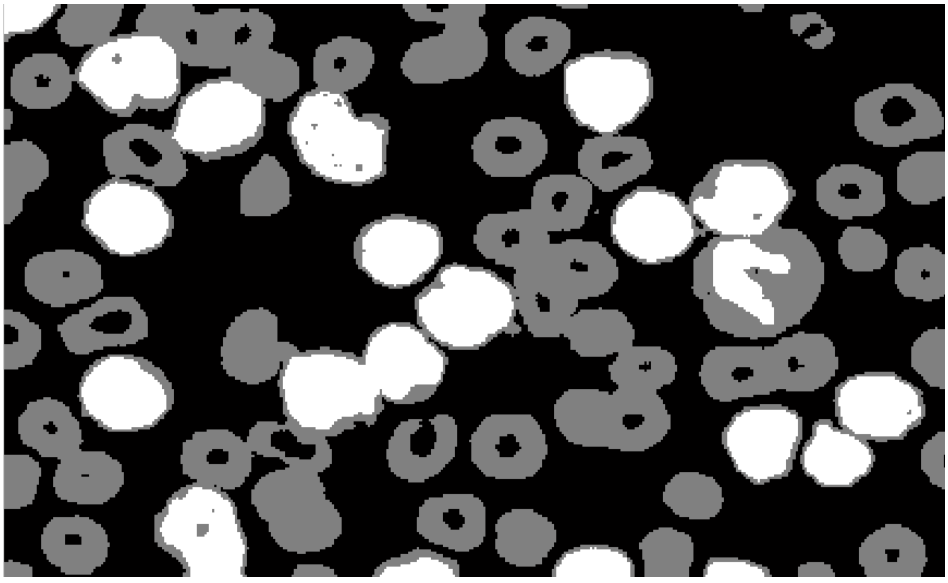


**Figure 5.3:** Cluster index image

Figure 5.3 shows the cluster index image which is important for applying the K-means algorithm. Cluster index image defines cluster to the each component of the image.

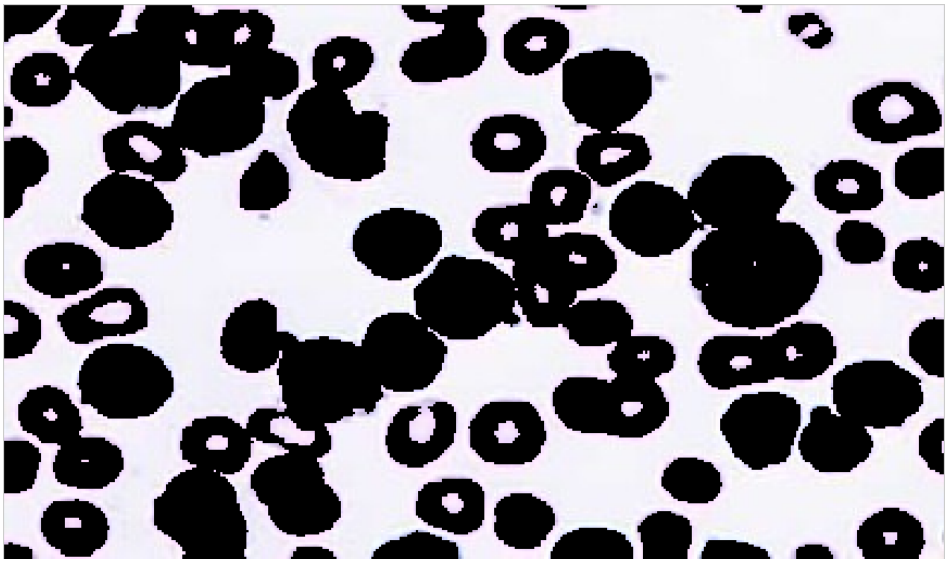


**Figure 5.4:** Background of image

The K-means clustering is applied on the figure 5.3 and we get three clusters: Background, Red blood cells cluster and white blood cells cluster. Figure 5.4 shows the background of the image.

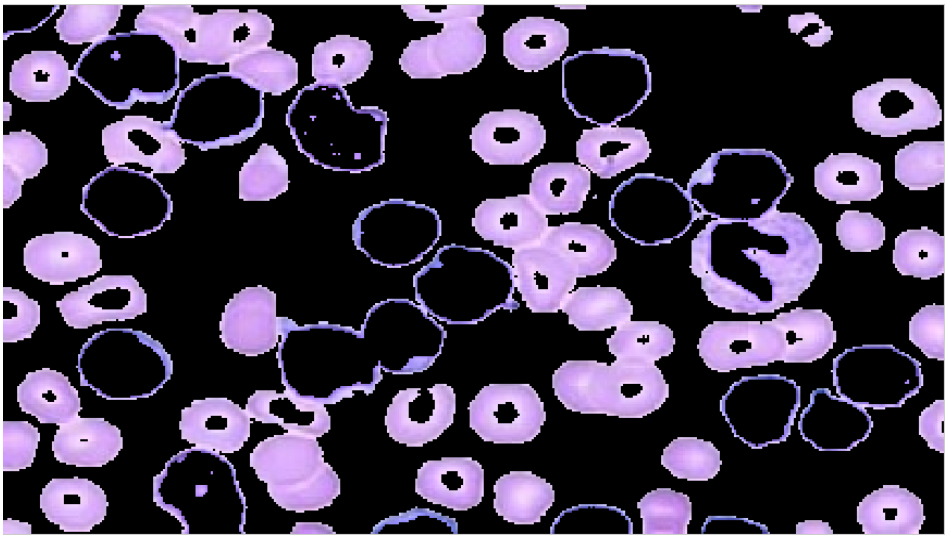


**Figure 5.5:** Red blood cells cluster

Figure 5 shows the Red blood cells cluster only by showing the red colored part as it is and making the remaining portion black.

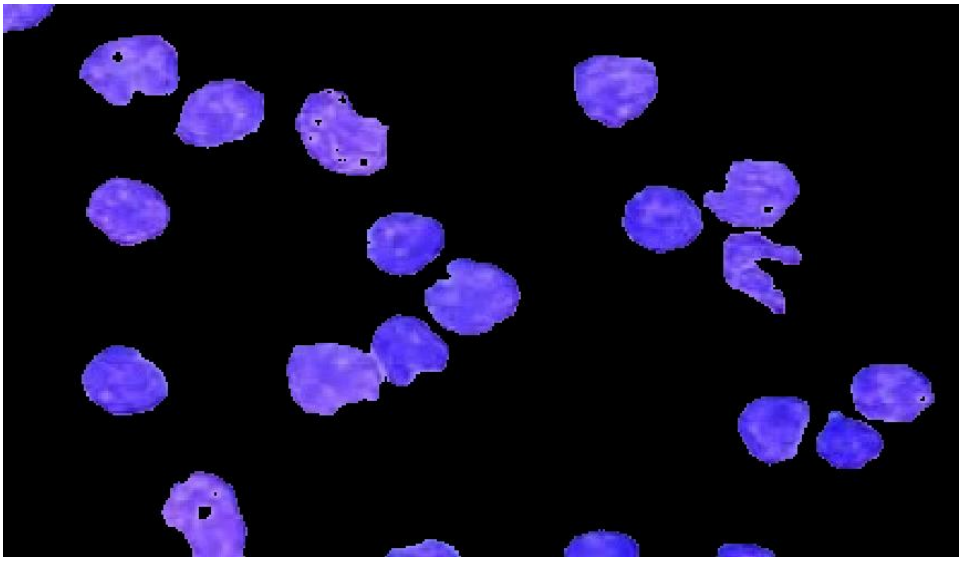


**Figure 5.6:** White blood cells cluster

Figure 5.6 shows the White blood cells cluster only by showing the blue colored part which is the nucleus of the cells.

The proposed system is tested by the different microscopic images and the accuracy is also calculated. The image dataset we have used here is ALL-IDB, proposed by Donida Labati [12]. The images are different in many terms of lightening, magnification and resolution. Thirty-three images are fetched from the whole ALL-IDB dataset which are taken from the same camera and same lightening conditions. The proposed system also shows the percentage of the infection present in the blood image.

**Table 5.1:** Proposed method performance for white blood cells identification

| Image | Manual Count | Auto Count | Accuracy |
|---|---|---|---|
| Image1 | 12 | 11 | 91% |
| Image2 | 9 | 5 | 55% |
| Image3 | 7 | 4 | 57% |
| Image4 | 8 | 8 | 100% |
| Image5 | 24 | 19 | 79% |
| Image6 | 18 | 18 | 100% |
| Image7 | 7 | 7 | 100% |
| Image8 | 17 | 16 | 94% |
| Image9 | 7 | 7 | 100% |
| Image10 | 12 | 12 | 100% |
| Image11 | 15 | 12 | 80% |
| Image12 | 12 | 12 | 100% |
| Image13 | 10 | 7 | 70% |
| Image14 | 5 | 3 | 60% |
| Image15 | 17 | 17 | 100% |
| Image16 | 16 | 16 | 100% |
| Image17 | 3 | 3 | 100% |
| Image18 | 8 | 8 | 100% |
| Image19 | 12 | 12 | 100% |

| | | | |
|---|---|---|---|
| Image20 | 2 | 2 | 100% |
| Image21 | 3 | 3 | 100% |
| Image22 | 5 | 5 | 100% |
| Image23 | 6 | 6 | 100% |
| Image24 | 4 | 4 | 100% |
| Image25 | 3 | 3 | 100% |
| Image26 | 5 | 5 | 100% |
| Image27 | 3 | 3 | 100% |
| Image28 | 2 | 2 | 100% |
| Image29 | 4 | 4 | 100% |
| Image30 | 3 | 3 | 100% |
| Image31 | 2 | 2 | 100% |
| Image32 | 2 | 2 | 100% |
| Image33 | 2 | 2 | 100% |

# Chapter 6

# Conclusion and Future work

The main focus of this thesis is to propose an automated system which can detect the leukemia from the microscopic image to improve the accuracy and reduce the time to detect than the manual approach. So many lives can be saved by using the proposed automated approach of leukemia detection.

## 6.1 Conclusion

The major part of this work is to segment the lymphocytes and myelocytes white blood cells for leukemia detection. The first phase of the proposed system is dealing with the image cleaning and noise removal for making the image ready for the further and accurate study. The second and major phase is the leucocytes identification from the image. The third phase is dealing with the nucleus and cytoplasm extraction from the image which can finally be used for the feature extraction in the last phase of the proposed system. This model has been tested against 33 images taken under same lightening condition and the accuracy achieved is 93.57%. We can also use the proposed system to find out the percentage of leukemia infection in microscopic image. We hope this approach will be beneficial for today's fast life and early detection of leukemia without any need of costly tests and with a better accuracy.

## 6.2 Future Scope

There are so many ways to make this system better in future. We can improve the segmentation scheme which can segment the overlapped cells also. There were found the use of multiple classifiers in some systems. We can also use multiple classifiers to improve the accuracy of the classification. Doing so will increase the cost but accuracy will also be improved. We can use parallel algorithm for the execution so that the execution time can be decreased.

# References

[1] Bandini, S., Vanneschi, L., Wuensche, A. and Shehata, A. B., (2009). Cellular Automata Pattern Recognition and Rule Evolution Through a Neuro-Genetic Approach. *Journal of Cellular Automata 4*. 171-181.

[2] Aguilera, D G. and Lahoz, J. G., (2008). Learning from Educational Software in 3d Cartography. *British Journal of Educational Technology 39.4.* 726 - 731.

[3] Davidson, R. and Harel, D., (1996). Drawing Graphs Nicely Using Simulated Annealing. *ACM Transactions on Graphics. Vol 15. No. 4*. 301 – 331.

[4] Golchin F and Paliwal, K. K., (2003). Quadtree-based classification in subband image coding. *Elsevier - Digital Signal Processing 13*. 656-668.

[5] Guieb, E. C., and Samaneigo. J. M., (2007). Image Noise Reduction Using Cellular Automata. *CMSC 190 Special Problem, Institute Of Computer Science*.

[6] Madhloom HT, Kareem SA, Ariffin H, Zaidan AA, Alanazi HO, Zaidan BB. Anautomated white blood cell nucleus localization and segmentation using imagearithmetic and automated threshold. J Appl Sci 2010;10(11):959–66.

[7] Kovalev VA, Grigoriev AY, Ahn H. Robust recognition of white blood cell images.In: Kavanaugh ME, Werner B, editors. Proceedings of the 13th international conference on pattern recognition, August 25–29. Vienna, Austria: IEEE Publisher;1996. p. 371–5.

[8] Scotti F. Robust segmentation and measurements techniques of white cells inblood microscope images. In: Daponte P, Linnenbrink T, editors. Proceedingsof the IEEE instrumentation and measurement technology conference, 24–27April. Sorrento, Italy: IEEE Publisher; 2006. p. 43–8.

[9] Piuri V, Scotti F. Morphological classification of blood leucocytes by microscope images. In: Proceedings of the IEEE international conference on computational intelligence for measurement systems and applications, 14–16 July. Boston,MA, USA: IEEE Publisher; 2004. p. 103–8.

[10] Halim NHA, Mashor MY, Hassan R. Automatic blasts counting for acute leukemia based on blood samples. Int J Res Rev Comput Sci 2011;2(August(4)).

[11] Mohapatra S, Patra D, Satpathy S. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. J NeuralComput Appl 2014;24:1887–904.

[12] Donida Labati R, Piuri V, Scotti F. ALL-IDB: the acute lymphoblastic leukemiaimage DataBase for image processing. In: Macq Benot, Schelkens Peter, editors.Proceedings of the 18th IEEE ICIP international conference on image processing,September 11–14. Brussels, Belgium: IEEE Publisher; 2011. p. 2045–8.

[13] Q. Liao and Y. Deng. An accurate segmentation method for white blood cell images. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pages 245 – 248, 2002.

[14] J. Angulo and G. Flandrin. Microscopic image analysis using mathematical morphology: Application to haematological cytology. In A. Mendez-Vilas, editor, *Science, Technology and Education of Microscopy: An overview*, pages 304–312. FORMATEX, Badajoz, Spain, 2003.

[15] N. Sinha and A.G. Ramakrishnan. Blood cell segmentation using em algorithm. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, 2002.

[16] N. T. Umpon. Patch based white blood cell nucleus segmentation using fuzzy clustering. *ECTI Transaction Electrical Electronics Communications*, 3(1):5–10, 2005.

[17] L.B. Dorini, R. Minetto, and N.J. Leite. White blood cell segmentation using morphological operators and scale-space analysis. In *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing*, pages 294–304, October 2007.

[18] D. Comaniciu and P. Meer. Cell image segmentation for diagnostic pathology. *Advanced Algorithm Approaches to Medical Image Segmentation: State-Of-The-Art Application in Cardiology, Neurology, Mammography and Pathology*, pages 541–558, 2001.

[19] L. Yang, P. Meer, and D.J. Foran. Unsupervised segmentation based on robust estimation and color active contour models. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):475 –486, September 2005.

[20] F. Yi, Z. Chongxun, P. Chen, and L. Li. White blood cell image segmentation using on-line trained neural network. In *Proceedings of the 27th Annual International*

*Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6476–6479, January 2005.

[21] M. Ghosh, D. Das, C. Chakraborty, and A. K. Ray. Automated leukocyte recognition using fuzzy divergence. *Micron*, 41(7):840–846, 2010.

[22] W. Shitong and W. Min. A new detection algorithm based on fuzzy cellular neural networks for white blood cell detection. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):5 –10, January 2006.

[23] S. Chinwaraphat, A. Sanpanich, C. Pintavirooj, M. Sangworasil, and P. Tosranon. A modified fuzzy clustering for white blood cell segmentation. In *Proceedings of the Third International Symposium on Biomedical Engineering*, volume 6, pages 2259–2261, 2008.

[24] C. Meurie, O. Lezoray, C. Charrier, and E. Elmoataz. Combination of multiple pixel classifiers for microscopic image segmentation. *IASTED International Journal of Robotics and Automation*, 20(2):63 – 69, 2005. Special Issue on Colour Image Processing and Analysis for Machine Vision.

[25] M. Ghosh, D. Das, C. Chakraborty, M.Pala, A.K. Maity, S.K. Pal, and A. K. Ray. Statistical pattern analysis of white blood cell nuclei morphometry. In *Proceedings of the IEEE Students Technology Symposium*, pages 59–66, April 2010.

[26] B. C. Ko, J. Gim, and J. Nam. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron*, 42(7):695 – 705, 2011.

[27] S. Serbouti, A. Duhamel, H. Harms, U. Gunzer, U.M. Aus, J.Y. Mary, and R. Beuscart. Image segmentation and classification methods to detect leukemias. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 13, pages 260–261, 1991.

[28] D.J. Foran, D. Comaniciu, P. Meer, and L.A. Goodell. Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *IEEE Transactions on Information Technology in Biomedicine*, 4(4):265–273, 2000.

[29] F. Scotti. Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images. In *Proceedings of IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, pages 96 – 101, July 2005.

[30] T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczynski. Automatic recognition of the blood cells of myelogenous leukemia using svm. In *Proceedings of IEEE International Joint Conference on Neural Networks*, volume 4, pages 2496 –2501, August 2005.

[31] Hazwani Abd Halim, Mohd Yusoff Mashor, and Rosline Hassan. Automatic blasts counting for acute leukemia based on blood samples. *International Journal of Research and Reviews in Computer Science*, 2(4):971–976, 2011.

[32] R. Seshadri, R. L. Jarvis, O. Jamal, and J. M. Skinner. A morphometric classification of acute lymphoblastic leukemia in children. *Medical and Pediatric Oncology*, 13(4):214–220, 1985.

[33] J. Angulo, J. Serra, and G. Flandrin. Quantitative descriptors of the lymphocytes. In *Proceedings of the 7th Congress of the European Society for Analytical Cellular Pathology*, pages 69–70, France, April 2001.

[34] J. Angulo, J. Klossa, and G. Flandrin. Ontology based lymphocyte population description using mathematical morphology on colour blood images. *Cellular and Molecular Biology*, 52(6):2 – 15, 2006.

[35] L. Gupta, S. Jayavanth, and A. Ramaiah. Identification of different types of lymphoblasts in acute lymphoblastic leukemia using relevance vector machines. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, USA, 2009.

[36] H. J. Escalante,M.M. Gomez, J. A. Gonzalez, P. G. Gil, L. Altamirano, C. A. Reyes, C. Reta, and A. Rosales. Acute leukemia classification by ensemble particle swarm model selection. *Artificial Intelligence in Medicine*, 55(3):163 – 175, 2012.

# List of Publications

1. Nimesh Patel and Ashutosh Mishra, "Automated Leukemia Detection using Microscopic Images", published at Second International Symposium on Computer Vision and The Internet (VISIONNET'15) Workshop on Signal Processing, Image Processing and Pattern Recognition (SIPR-2015), Kochi, Kerala, India, IEEE, 2015.

# Video Link

http://youtu.be/N32lhwp0YbJ