Trinion Based WBC Segmentation Using Texture to Detect Acute Lymphoblastic Leukemia

By

AMIN HUSSEIN

A Thesis Submitted in Partial Fulfillment of the Requirement of the Degree of Master of Science in Biomedical Engineering

Advisor: Dawit Assefa Haile (PhD)

October 2021

Declaration

I, the undersigned, declare that this thesis is my original work. It has never been presented for a degree in any other institution and that all sources of materials used in it have been duly acknowledged.

Name ……..

Signature …….

Date ……

This MSc thesis has been submitted for examination with my approval as his advisor.

…….

Dawit Assfa Haile (PhD)

Addis Ababa University

School of Graduate Studies

Certificate of Examination

This is certify that the thesis prepared by Amin Hussein entitled *"Trinion Based WBC Segmentation Using Texture to detect Acute Lymphoblastic Leukemia"* submitted in partial fulfillment of the requirements for the degree of Master of Science in Biomedical Engineering complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Examiner ……Signature……..Date……

Examiner…….Signature……..Date……..…

Examiner ……Signature……..Date……..…

Adviser ……...Signature …….....Date…….

Abstract

Many diseases are detected based on examination of microscopic images of blood samples. Changes in the blood condition show the development of diseases in an individual. One type of disease caused by change of blood condition is Leukemia. Leukemia can cause early death when it is not treated on time. In Ethiopia, Leukemia accounts to about 35.5% of hematological admissions. The death rate in Ethiopia due to Leukemia is different with time and region. The average death rate is increasing from time to time and shows variation between country side and urban population. Reports from the World Health Organization (WHO) show that death rate due to Leukemia in Ethiopia has reached 5.56% and ranked 18th highest in the world.

Leukemia originates in the bone marrow, a thin material inside the bone. Leukemia is detected by analyzing white blood cells (WBCs also called Leucocytes), one of the constituents of blood along with red blood cells (RBC or Erythrocytes), platelets and blood plasma. WBCs have five different types (Lymphocytes, Myelocytes, Neutrophils, Basophils and Eosinophils) and among these Lymphocytes and Myelocytes are the ones that could start to change in the bone marrow and get infected and become Leukemic or infected cells. These Leukemia cells have strange properties compared to the normal cells in that their growth is abnormal and they survive much longer than the normal cells. They also interrupt functions of the normal cells. Through time, the normal cells perish while leukemia cells still survive. Old leukemia cells last for a longer time and production of new leukemia continue in an abnormal way.

Traditionally, Leukemia detection is carried out manually based on visual examination of microscopic images of blood samples. This is lengthy and time taking process which depends on the skills and experiences of the observer which makes the process subjective. In this regard, computer based automated schemes play their great role and several efforts have been made in the literature to develop such schemes.

In the current study, a novel mathematical technique for Leukemia detection based on holistic analysis of color microscopic images of blood samples is proposed. The approach utilizes a holistic representation of microscopic blood images in the three (Trinion) space and applies trinion based Fourier transform implemented in the L*a*b color space to extract useful higher order features to segment normal and infected WBCs and classify them. The technique has been applied in analyzing microscopic images acquired from standard ALL-IDB database. Classification of normal and Leukemic WBCs was performed based of Artificial Neural Network (ANN) which resulted in 95.7% sensitivity, 100% specificity and 97.6% accuracy.

Acknowledgment

First and foremost I am very grateful to ALLAH the Almighty for the blessings and guidance He has bestowed upon me throughout the entire period of my thesis undertaking. My deepest appreciation goes to my supervisor Dawit Assefa Haile, PhD, for his support and kindness in advising me to keep improving my knowledge and to keep believing in my abilities.

A similar level of gratitude is for the staff of the Center of Biomedical Engineering in AAU. It is unlikely that I would have reached completion without their encouragement and support. I express my appreciation to everyone involved directly and indirectly to the success of this thesis.

Last but not least, my family for their understanding, support, patience, and encouragement. Thank you for all the support, comments and guidance.

Table of Contents

Abstract…… III

Acknowledgement…… IV

Table of Contents…..…… V

List of Figures..……..……VII

List of Tables……..VIII

List of Abbreviations (Acronyms) …….....IX

Chapter one……....1

1. Introduction ……..….1

1.1 Background……....1

1.2 Problem Statement ……....2

1.3 Objective of the study ……..…...3

1.3.1 General Objective ……..….3

1.3.2 Specific Objectives ……3

1.4 Scope of the project ……....3

1.5 Thesis Contribution ……....3

1.6 Organization of the Thesis…….....4

Chapter Two ……...5

2. Blood and Blood Diseases……...5

2.1 Blood……..8

2.2 Blood Disease ……8

2.3 Blood Cancer **(**Hematological Malignancies)……..8

2.4 Leukemia…….……9

2.4.1.1 Acute Lymphoblastic leukemia (ALL)……10

2.4.1.1.1 Etiology ……...11

2.4.1.1.2 Clinical Signs and Symptoms ……12

2.4.1.1.3 Convectional Diagnosis Techniques ……...12

2.4. Limitation of Convectional Techniques…….……...14

Chapter Three ……....…15

3. Hematological Image Processing ……... 15

3.1.History of Digital Image Processing In Medical Applications……15

3.2. Image Formation ……...……..…15

3.3. Color and Color space conversation ……..16

3.4. Color Image Segmentation …….....18
3.5. Feature Extraction ……..21
3.6. Classification …….....23
3.7. Performance Analysis…….…….25

Chapter Four ……..26

4. Proposed Trinion Based WBC Segmentation ……...…….…..….26
4.1.General Overview on the Proposed Method……....26
4.2.Image Acquisition …….....26
4.3.Image Pre-Processing ……..27
4.4.TFT Based WBCs Segmentation Using Textural Features ……..……..28
4.4.1. Data Dimensionality Reduction Using PCA ……...29
4.4.2. Textural Features Extraction and Selection of Signature map…….....30
4.5.Image Cleaning from the Signature Map……...……..32
4.6.Geometric Feature Extraction from the Signature Map……...32
4.7.Classification Using Artificial Neural Network (ANN)...…….....33
4.8.Classification Performance Measurement ……..35

Chapter Five ……37

5. Result and Discussion …….……..35
5.1. Detection of WBCs from Microscopic Blood Images……35
5.2. Signature Map Generation ……36
5.3.Image Cleaning on the Signature Maps ……...……...37
5.4. Extraction of Geometric Features……...39
5.5. Classification using ANN ……...……...40
5.5.1. Morphological level classification ……..…41

Chapter Six ……..47

6. Conclusion and Recommendation …….…..47
6.1.Conclusion …….....47
6.2.Recommendation……48

Reference ……,……...49

Appendix ……..55

List of Figures

Figure 2.1: The major components of whole blood…….…….…….5

Figure 2.2: Blood cells formation in the bone marrow …….…….6

Figure 2.3: Developmental pathway of Leukocyte ……..…7

Figure 2.4: Common symptoms of leukemia …….…...13

Figure 2.5: General steps that hematopathologists follow in the lab to diagnose leukemia …..…14

Figure 3.1: Image formation …….……16

Figure 3.2: Black and white leukemia image ……...……..…...16

Figure 3.3: Components of RGB leukemia image: (a) Red component, (b) Green component, (c) Blue component and (d) Original RGB image……..…17

Figure 4.1: Block diagram of the proposed method ……..……..25

Figure 4.2: (a) Examples of blast cells, (b) Examples of normal BCs…….…...……..…26

Figure 4.3: The general algorithmic steps followed for geometric features extraction..…….……..31

Figure 4.4: Structure of a human neuron [83].…….…..32

Figure 4.5: The basic architecture of a neural network……..…33

Figure 5.1: Unhealthy microscopic blood image (a) and healthy microscopic blood image in the RGB space; Unhealthy image (c) and healthy microscopic image in L*a*b color space...….35-36

Figure 5.2: Signature map generated based on cluster prominence feature on unhealthy image sample (a), signature map generated based on cluster prominence feature on healthy image (b). ……..37

Figure 5.3: General steps of cleaning the signature maps…….……...…….…...38

Figure 5.4: The final signature map of the infected WBCs (a) and healthy WBCs (b) after the morphological cleaning operation……..……..…39

Figure 5.5: Classifier performance graph…….…….....42

Figure 5.6: Neural network training error histogram plot.…….....42

Figure 5.7: Neural network training, testing and validation performance…….....43

Figure 5.8: ROC curves for the neural network training, testing and validation states….....44

List of Tables

Table 2.1: Types of Leukocytes ……..…7

Table 2.2: Examples of blood diseases……....8

Table 2.3: Classification of leukemia based on clinical course and origin of leukemia cell population..……..…….……...……9

Table 2.4: Classification of ALLs based on the FAB scheme..……...…10-11

Table 4.1: Formulas used to compute Haralick textural features [82]……..…..29-30

Table 4.2: Confusion matrix for two classes …...…….…….34

Table 5.1: Geometrical features extracted from selected healthy segmented WBCs……..…….40

Table 5.2: Geometrical features extracted from ALL infected segmented WBCs…..……..……40

Table 5.3: Computed sensitivity, specificity and overall accuracy matrices for the proposed scheme……44

List of Abbreviations

ANN	Artificial Neural Network
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
CLL	Chronic Lymphoblastic Leukemia
CML	Chronic Myeloid Leukemia
CT	Computed Tomography
FAB	French-American-British
FN	False Negative
FP	False Positive
GDA	Generalized Discriminate Analysis
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
MLP	Multilayer Perception
MSE	Mean Square Error
NBC	Naive Bayesin Classification
NK	Natural Killer
NMF	Negative Matrix Factorization
PCA	Principal Component Analysis
RBC	Red Blood Cells
RBFN	Radial Basis Function Network
SVM	Supper Vector Machine
TFT	Trinion Fourier Transform
TN	True Negative
TP	True Positive
WBC	White Blood Cells
WHO	World Health Organization

Chapter One

1. Introduction

1.1Background

The term disease implies discomfort, or absence of ease within the body. Whenever the normal functioning of the body or any of its part becomes impaired, diseases occur and may require medical treatment [1]. In general, diseases can be classified on the basis of their cause and cell of origin as infectious, immunological, endocrine, genetic, neoplastic, traumatic and the like. Physicians across the globe are interested in understanding the biology of a disease and how it can be prevented or treated [2]. Among all diseases, the quest for understanding cancer is quite high. All cancers are characterized by uncontrolled growth of abnormal cells, invade surrounding tissues, metastasize (spread to distant sites), and eventually killing the host where it originates [3]. Cancer can develop in individuals of any race, gender, age, socioeconomic status, or culture and can involve any type of cells, tissues or organs of the human body. Cancers can be of two forms: a solid mass (tumors) or liquid cancer. As a cancer which develops from cells in the blood, bone marrow, and lymphatic system, leukemia is one type of liquid cancer.

Leukemia cells are also referred to as blast cells. Leukemia is a type of blood cancer which was first recognized in 1845 by Bennett in Scotland and Virchow in Germany [4]. In their first two patients, it was the post-mortem appearance of the blood which first gave the hint of an abnormal condition. In Virchow's case, the blood vessels contained a "yellowish-white almost greenish mass" [5]. Microscopically, there were a few red blood corpuscles and some colorless white bodies, which are found in the blood of a normal person. The relationship between the red and the colorless corpuscles was the reverse of the normal ratio. Virchow introduced the term 'Leukemia' to describe the condition [6].

Leukemia is a disease of unknown cause where the bone marrow produces a large number of abnormal cells. Each bone contains a thin material inside it which is also known as a bone marrow. The components of blood are RBC (erythrocytes), WBC (leucocytes), platelets and plasma. Leukemia is detected only by analyzing the WBCs. WBCs have five different types (lymphocytes, myelocytes, neutrophil, basophil and eosinophil) and among these lymphocytes and myelocytes start to change in the bone marrow and they get infected and become leukemic or infected cells. These leukemia cells show strange properties than normal cells. Their growth is abnormal and survival time is much more than that of the normal cells. They interrupt normal cells not to carry out their task. Through time, the normal cells perish while leukemia cells still survive. Old leukemia cells last for a longer time and production of new leukemia continue in an abnormal way.

There are four general types of leukemia, namely Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphoblastic Leukemia (CLL), and Chronic Myeloid

Leukemia (CML). Traditionally, the process of detection and classification is manual, based on the use of different techniques. The commonest is microscope examination in which case diagnosis takes up to few days. The diagnosis generally depends on the expert's experience and judgment. Traditionally, Leukemia detection is carried out manually based on visual examination of microscopic images of blood samples. This is lengthy and time taking process which depends on the skills and experiences of the observer which makes the process subjective. In this regard, computer based automated schemes play their great role and several efforts have been made in the literature to develop such schemes. The motivation behind the current thesis project is to improve the diagnosis process by automating it and reducing the time it takes.

Nowadays, medical image processing is one of the fastest growing fields in medicine, both in clinical settings and research. Imaging improves patient care, contributing to areas such as personalized medicine with individually tailored treatment, increasing evidence-based decision making within health care, reducing complications during and after surgery, and allowing better understanding of the effects of treatments in various diseases [7]. Image processing in medical research is becoming a subject of prime focus due to its tremendous potential for the public health sector and the scientific community in general. In particular, imaging applications are emerging as a new opportunity for innovation at the meeting point between medicine and computer science. Many software companies and research groups focus on the development of image processing applications for the medical world, for example to improve quality of low-resolution photographic images and produce effective high-quality imagery [8]. Collaboration with clinicians has allowed the extraction of useful information contributing to more efficient diagnosis, especially in the treatment and study of cancer. In most people's minds, there is no more frightening disease than cancer, often viewed as untreatable, unbearable, and painful disease with no cure. Indeed it is a serious, potentially life-threatening illness [9].

The current research focuses on ALL, a serious illness caused by the abnormal growth and development of WBCs. It starts in the bone marrow and move thought the peripheral blood. As the blast cells build up, they hamper the body's natural ability to fight infection and stop bleeding. Therefore, the disease requires immediate treatments.

1.2 Problem Statement

There are many techniques that have been developed in the literature for automated detection of Leukemia cells from microscopic blood images but people still use manual methods like physical examination, CBC, blood chemistry studies and bone marrow aspiration particularly in low resource settings as it is often the case in developing world. The manual detection techniques of Leukemia are prone to various drawbacks. To circumvent this, several algorithmic techniques have been suggested in the literature. Most use separate analysis of the three channels of the color microscopic images of blood samples. Others tried to use different color spaces for the task. Morphological schemes, integral transforms such as the Gabor transform and wavelets and the like have been proposed previously for more effective automated detection of the Leukemia

cells. It is the intent of this thesis to review commonest automated techniques proposed previously in the literature and discuss their drawbacks. In this study, a novel mathematical technique for analysis of color microscopic blood images has been proposed. The approach follows a holistic representation of microscopic blood images in the three (trinion) space and applies trinion based WBCs segmentation by extracting higher order features in the L*a*b color space. The segmentation process is followed by morphological cleaning operations. Classification of normal and Leukemic WBCs was performed based of Artificial Neural Network (ANN).

1.3 Objective of the Study

1.3.1 General Objective

The main goal of this thesis is to establish effective and efficient technique to automate Leukemia detection based on analysis of microscopic blood sample images.

1.3.2 Specific Objectives

- Identify best color space for use in effective segmentation of Leukemia cells from the RGB microscopic images available;
- Implement the necessary pre-processing algorithms on the given microscopic blood images to remove noise and other unwanted signals;
- Implement an effective mapping of the converted color images using vectorial concepts using principles of Clifford algebra;
- Implement feature extraction procedure for effective classification/clustering of Leukemia cells and fully segment the images;
- Perform qualitative and quantitative analysis to evaluate the performance of the proposed image processing scheme against available ground truth information and identify promising clinical implications.

1.4 Scope of the Project

This work is only at the testing stage and some of the investigation will only be preliminary particularly regarding its clinical significance. This might require extensive observer study and implementation of the scheme on various data sets. Moreover, only ALL type Leukemia cases are considered and discussion on the applications of the proposed method on the other three types of Leukemia is beyond the scope of the current work.

1.5 Thesis Contribution

The major contributions of the thesis are:

- A new, texture based segmentation scheme developed in the trinion space has been presented for healthy as well as leukemic WBCs based from microscopic blood sample images.
- Combination of textural and morphological features has been used to classify mature lymphocytes and lymphoblasts using ANN classifier.

1.6 Organization of the Thesis

The rest of the thesis is organized in the following manner. Chapter 2 presents the detail information on the blood and blood related diseases in the circulation system. The most common hematological disorders are also presented in the chapter with its traditional diagnostic techniques. Chapter 3 discusses overview of basic concepts in digital image processing, color image processing and related issues including literature reviews on different segmentation techniques for the application of leukemia detection. Chapter 4 presents the proposed methodology for use in holistically analyzing the microscopic blood images for the purpose of detection of ALL. This chapter incorporates a way of selecting suitable color space and its transformation, a holistic representation of the color images in the three (trinion) space, a scheme for extracting second order statistical features for the purpose of generating signature maps and morphological analysis for further classifying the segmented images. The performance evaluation scheme based on different accuracy measurements is also presented. Chapter 5 presents and discusses the results obtained by the proposed method for accurate detection of ALL. Chapter 6 presents a summary of the accomplishments of the thesis along with possible recommendations for future work.

Chapter Two

2. Blood and Blood Disease

2.1 Blood

Blood is a fluid connective tissue which circulates through the heart and blood vessels. It appears to be a thick, homogeneous liquid, but the microscope reveals that it has both cellular and liquid components. Blood is a specialized connective tissue in which living blood cells, called the formed elements, are suspended in a nonliving fluid matrix called plasma. If we spin a sample of blood in a centrifuge, centrifugal force packs down the heavier formed elements and the less dense plasma remains at the top as shown Fig. 2.1. Most of the reddish mass at the bottom of the tube contains red blood cells that transport oxygen. A thin, whitish layer called the buffy coat is present at the erythrocyte plasma junction. This layer contains leukocytes, the WBCs that act in various ways to protect the body, and platelets, cell fragments that help stop bleeding [10].

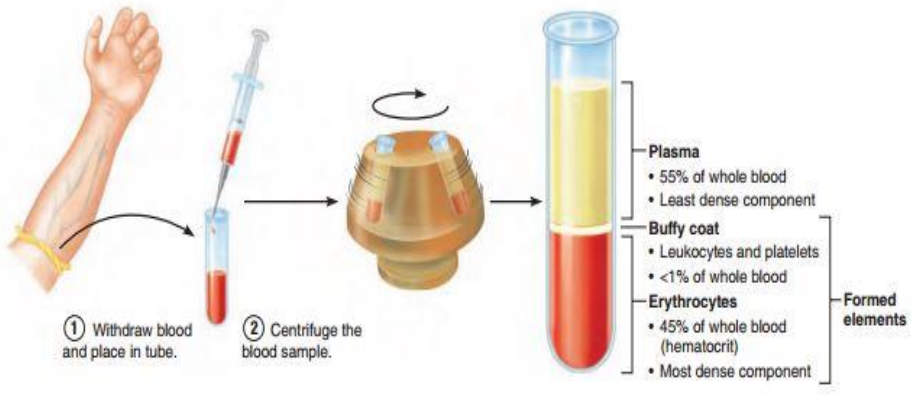


Figure 2.1: The major component of whole blood [10].

Blood performs a number of functions. It transports oxygen from the lungs and nutrients from the digestive tract to the tissues, transports metabolic waste products from the cells to elimination sites (to the lungs to eliminate carbon dioxide, to the kidneys to dispose of nitrogenous wastes in urine), and transports hormones from the endocrine organs to their target organs. Blood maintains appropriate body temperature by absorbing and distributing heat throughout the body and to the skin surface to encourage heat loss, maintaining normal pH in body tissues such as many blood proteins and other blood borne solutes act as buffers to prevent excessive or abrupt changes in blood pH that could jeopardize normal cell activities, and maintaining adequate fluid volume in the circulatory system. Preventing blood loss when a blood vessel is damaged, platelets and plasma protein initiate clot formation, halting blood loss, and preventing infection, drifting along in blood as antibodies, complement proteins and WBCs, all of which help defend the body against foreign invaders such as bacteria and viruses [10].

Blood is composed of different types of cells suspended in a pale yellow colored transparent fluid called plasma [11]. There are three types of blood cells:

a. Red Blood Cell (Erythrocyte): transport oxygen in which it combines oxygen in the lungs and carries it to tissues where it is needed for metabolic processes (the release of energy by burning food).
b. White Blood Cell (Leukocyte): protects the body from microbes and other foreign materials that come from external environment and help in the immune process.
c. Platelet (Thrombocyte): contains a variety of substance that promotes blood clotting.

The process of blood cell formation is known as hematopoiesis (hemato means blood, poiesis means to make) and takes place in the bone marrow. Initially all blood cells originate from pluripotent stem cells and undergo several developmental stages before distinct cells of each type are formed (as shown the Fig. 2.2) and finally enter the peripheral blood stream.

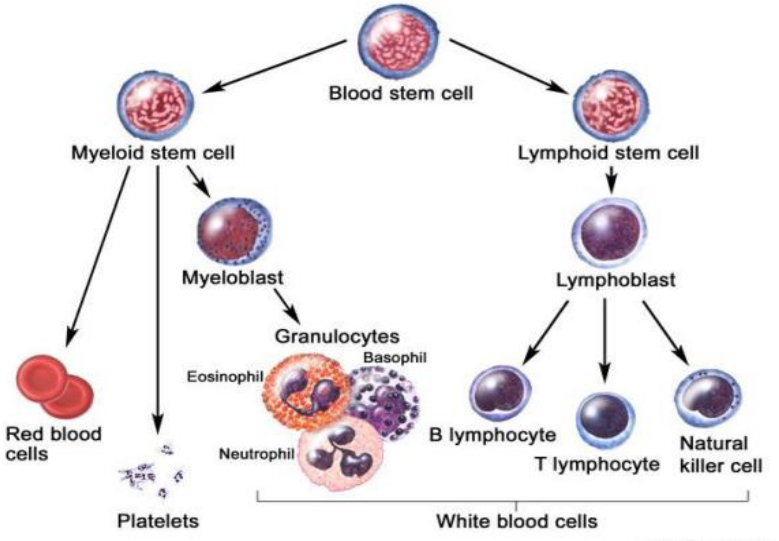


Figure 2.2: Blood cells formation in the bone marrow [10].

White blood cells (WBCs) are responsible for defending the body against infections caused by microbes and other foreign materials. They are the largest blood cells and account for about 1% of the blood volume. Unlike erythrocytes, leukocytes have a nuclei and each cell is made up of a nucleus and cytoplasm. The nucleus contains chromatin material and is chemically deoxyribonucleic acid (DNA) carrying genetic messages. Normally, human peripheral blood contains mature leukocytes and can be classified into two major groups of cells: polymorph nuclear leukocytes (granulocytes include basophile, neutrophils, and eosinophils) and mononuclear leukocytes (agranulocytes include lymphocytes and monocytes) [12]. This classification is based on nucleus morphology and presence of cytoplasmic granules. Table 2.1 presents the classification of leukocytes and its percentage composition.

Table 2.1: Types of Leukocytes.

No	Major types	Specific types	Percentage of WBC
1	Granulocytes	Neutrophils	50%-70%
		Eosinophils	Less than 5%
		Basophils	Less than 1%
2	Agranulocytes	Lymphocytes	25%-35%
		Monocytes	4%-10%

Lymphocytes are further subdivided into B–lymphocytes, which are synthesized in the bone marrow, T–lymphocytes from the thymus gland and natural killer (NK) cells. They continuously circulate between tissues and blood stream and are accountable for body's immune responses.

Monocytes are large mononuclear cells that originate in the red bone marrow and spleen. They are phagocytic in nature and are part of body's defense mechanism against bacterial and fungal infections. Monocytes are also responsible for the cleaning of dying body cells and additionally, immature leukocytes, i.e. unsegmented neutrophils, myelocytes, metamyelocytes, promyelocytes, myeloblasts, monoblasts. Lymphoblasts are also present in human body and are normally found in the bone marrow. But in individuals with unregulated or increased growth, they get spilled to peripheral blood and different types of WBC disorder are observed. Figure 2.3 present the general developmental pathway of leukocyte formation from hematopoietic stem cells.

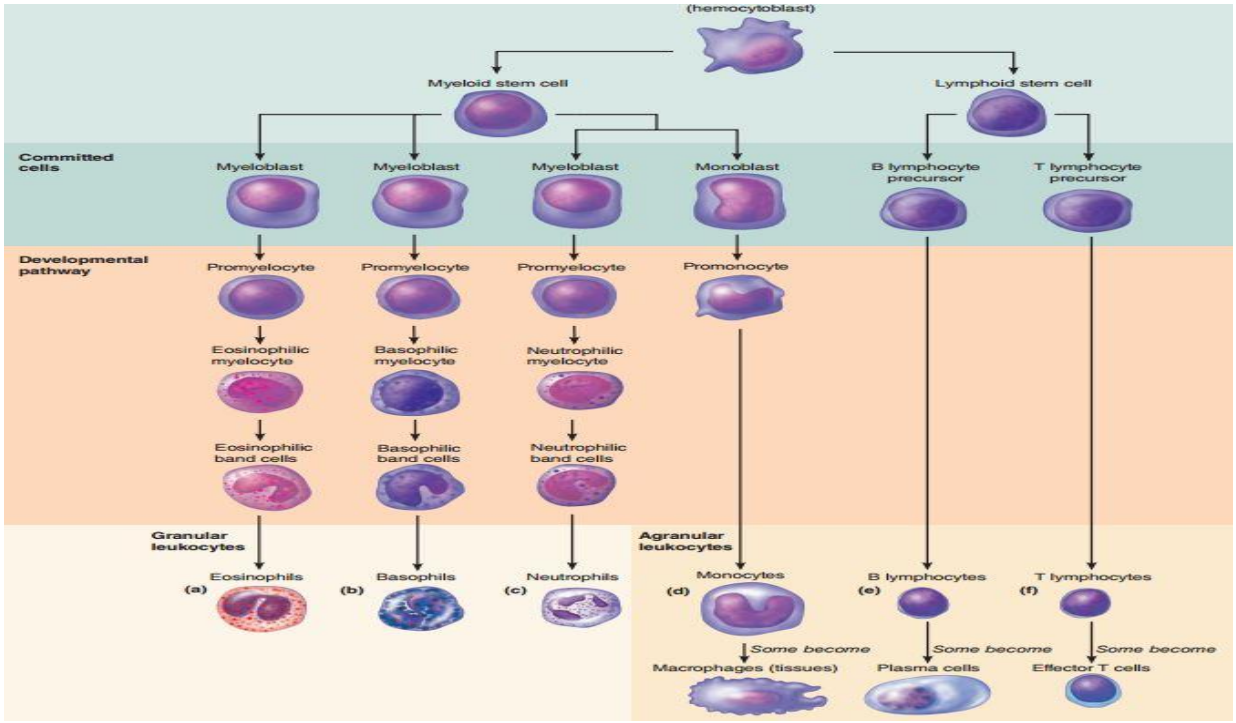


Figure 2.3: Developmental pathway of Leukocyte [10].

2.2. Blood Diseases

Hematology is the study of blood diseases and is diagnosed by medical experts known as hematopathologists. Hematological disorders can be classified in three ways: by the types of blood cells which are affected, according to functional disorders of the blood and lymphoid organs, and neoplastic disorders of the blood and lymphoid organs [13]. Moreover the neoplastic diseases can also be further classified into: malignant disorders which are conditions that involve the transformation of cell morphology, textures or colors, and nonmalignant disorders which are conditions that involve increase or decrease in the amount of cells.

Table 2.2 lists few examples of blood diseases along with the basic pathology they belong to. In the current study only malignant disorder of leukocytes is considered and a brief introduction on hematological malignancies is presented in the following section.

Table 2.2: Examples of blood diseases.

No	Disorder	Pathology	Disease
1	Erythrocyte	Increase RBC	Polycythemia
		Decrease RBC	Anemia
2	Leukocyte	Increase WBC	Eosinophilia infection of mononucleosis sepsis
		Decrease WBC	Leukopenia
		Malignant disorders of WBC	Leukemia Lymphomas
3	Hemostatic	Quantitative platelet disorder	Primary thrombocythemia Allergic Purpura
		Coagulation Disorder	Hemophilia
		Vascular Disorder	Purpura Simplex

`

2.3 Blood Cancer (Hematological Malignancies)

Our bodies are built up by cells which in normal way reproduce and grow in a certain limit under body controlled mechanism. However, this process sometimes goes wrong and cells reproduce and grow in abnormal way with both external factors (tobacco, infections organisms, chemicals, and radiation) and internal factors (inherited mutation, hormones, immune conditions, and mutations that occur from metabolism). These abnormal lesions can be anywhere in the body (such as blood, lymph node, bone, breast, skin, colon, or nerve tissue), and normally called growth or mass of tissue, called tumor. These abnormal lesions are called cancers when they are malignant. Some cancers, such as leukemia, do not form tumors. According to the American Cancer Society, among various types of human cancers, hematological malignancies account for the highest percentage of all cancers worldwide.

Hematological malignancies are heterogeneous group of cancers of the blood, bone marrow and lymph node. Such malignancies come from either of the two major blood cells: myeloid or

lymphoid [14]. Those diseases with myeloid origins are myelodysplastic syndromes, myelogenous leukemia and those with lymphoid origin are lymphomas, lymphocytic leukemia and myeloma. In Ethiopia, a total of 71 malignancy cases were found to be admitted to the pediatrics ward over 3 years from September 2010 to August 2013 and 43 (60.6 %) were found to have hematological malignancy (13 leukemia, 5 Hodgkin lymphoma, 3 non-Hodgkin lymphoma, 22 unspecified hematological malignancies in the registration) [15]. The high mortality rate of leukemia is mainly due to late diagnosis, and is mainly because of the symptoms of leukemia tend to mimic those of other common diseases. Due to unavailability of experienced pathologists and adequate laboratory facilities in district level hospitals of Ethiopia, many leukemia patients are initially misdiagnosed leading of patient deaths.

2.4 Leukemia

Leukemia is a liquid cancer which develops from cells in the blood, bone marrow, and lymphatic system. The basic difference of leukemia from other cancer types is that it does not form solid masses or tumors. The abnormal WBCs flow from the bone marrow to peripheral blood and disturb the normal functions of WBCs, RBCs and platelets. This can affect a patient in many ways: decrease in the amount of RBCs can result in anemia while drop in platelet count decreases the ability of clotting and the abnormal natures of WBCs limit the body's ability to fight infections. The symptoms of leukemia include fatigue, frequent infections, and easy bruising and bleeding. Depending on the clinical course, leukemia can be classified as either acute (with rapidly progressing disease with a predominance of highly immature blast cells) or chronic (which denotes slowly progressing disease with increased number of more mature cells) [16]. .Another classification of leukemia is developed to further identify differences in the response to treatment, prognosis and that is based on the hematopoietic cell of origin, i.e. myelocytic (myeloid) or lymphocytic (lymphoid). A basic classification of leukemia based on both clinical course and the origin of leukemia cell population is described in Table 2.3.

Table 2.3: Classification of leukemia based on clinical course and origin of leukemia cell population.

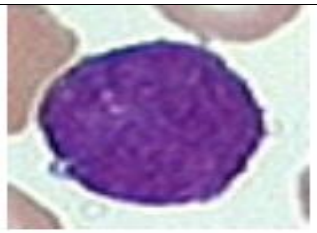
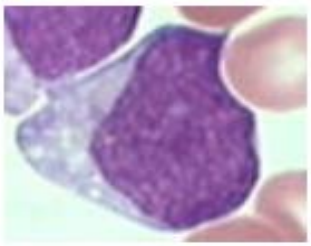
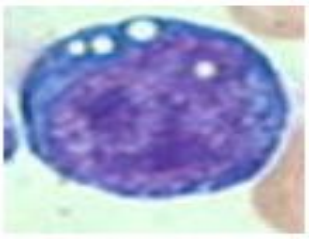
Clinical course	Origin of the cells	Types of leukemia	Description
Acute	Lymphoid	Acute Lymphoblastic Leukemia (ALL)	Common in children and affects adults over the age 65
	Myeloid	Acute myeloid Leukemia (AML)	Develop both in children and adult
Chronic	Lymphoid	Chronic Lymphoblastic Leukemia (CLL)	Common in adult over the age 55
	Myeloid	Chronic myeloid Leukemia (CML)	Mainly in adult, rare in children

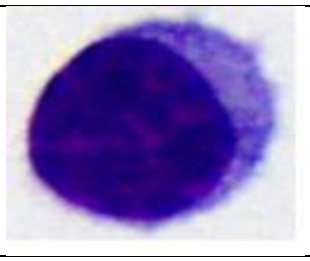
According to WHO, acute leukemia is defined as a malignant neoplasm with more than 20% blasts (myeloid or lymphoid) in the peripheral blood or bone marrow. In this thesis work, only one type of acute condition of malignant proliferation of lymphoid cells, ALL, is considered.

2.5 Acute Lymphoblastic Leukemia (ALL)

Acute lymphocytes leukemia (ALL) is a malignant disease caused by the genetic alteration of (lymphoid precursors proliferate and replace the normal hematopoietic cells of the marrow) the lymphocyte blast cells of the bone marrow. ALL is characterized by excessive production of lymphocytes in the bone marrow preventing normal hematopaiesis mechanism of the body and causes death, if not treated early. Clinically and biologically, features of ALL are sufficiently distinct from myeloid type of leukemia so that different diagnostics and treatment can be used. According to the French-American-British (FAB) scheme, ALLs could be classified into three based on their morphological variability (see Table 2.4).

Table 2.4: Classification of ALLs based on the FAB scheme.

FAB classification	Description	Typical Image Appearance
L1	Blasts are small and homogeneous, the nuclei are round and regular, and cytoplasm is scanty and usually without vacuoles	
L2	Blasts are large and heterogeneous, the nuclei are irregular, and the volume of cytoplasm is variable, but often abundant and may contain vacuoles.	
L3	Blasts are moderately large in size and homogeneous, the nuclei are regular and round oval in shape, the volume of cytoplasm is moderate and contains prominent vacuoles.	
Normal lymphocytes		
Description		Image

Lymphocytes are regular in shape having a compact nucleus with regular and continuous edges	

Globally, over 250,000 people are diagnosed with leukemia each year, accounting for 2.5% of all cancers. ALL accounts for approximately 80% of all leukemia patients and 30% of all cancers in children worldwide. Even though ALL is more prevalent in children and adolescents, it can appear in people of any age group and around 20% of adult acute leukemia cases are found to be ALL worldwide. Globally, incidence of ALL is found to be higher among males compared to females by nearly 40%, and the overall incidence of ALL in blacks is lower by 43% than in whites [17, 18].

2.5.1 Etiology

Cancer is a major societal burden worldwide. Even after years of research, surprisingly little is known about the exact cause of many cancers including leukemia. However, clinical evidences suggest that a variety of factors may be etiologically involved. Important etiological factors contributing to the development of ALL can be broadly classified as biological, physical and chemical factors [19]. Some of the evidences implicating chromosomal alterations, viruses, ionizing radiation and exposure to benzene in leukemogenesis are discussed below under biological, physical and chemical etiological factors.

A. Biological Factors

Cytogenetic Abnormalities: Hereditary syndromes are associated with cytogenetic abnormalities and have been linked to ALL [20]. These abnormalities include germ–line karyotype abnormalities, somatic karyotypic abnormalities, translocations, and deletions. The germ–line abnormalities associated with childhood leukemia include down syndrome and bloom syndrome. Somatic abnormalities are also associated with childhood leukemia. Translocations and deletions are also frequently found in ALL cases.

Infectious Etiology: Several lines of scientific evidences support the possibility that infections might cause ALL. The most widely accepted theory of causation of childhood ALL by infectious etiology was first proposed by Kinlen [21]. However, no specific virus or microbes are found to be associated with ALL.

B. Physical Factors

Ionizing Radiation: exposure to radiation in different forms has shown a strong and consistent association with ALL among children as well as adults. The most important evidence of ionizing radiation as an etiologic agent for ALL came from the studies of survivors of atomic bomb blasts in Japan [22]. There is also evidence for increased risk of ALL incidences in prenatal associated exposure to X–rays through radiography of pregnant women's abdomen [23].

Non ionizing Radiation: Epidemiological studies have also found positive association between ALL and residential exposure to electric and magnetic fields [24].

C. **Chemical Factors**

Solvents: Substantial number of epidemiologic studies have described elevated risks of childhood leukemia associated with parental occupational exposure to solvents, glues, exhausts, and paints [25]. Often, workers in various occupations, such as shoe, leather, rubber and printing industry are exposed to benzene and pose increased risk of leukemia.

Pesticides: Various hypotheses exist that suggest a link between ALL and pesticides [26]. Excessive use of organophosphates as pesticides on crops, fruits, and vegetables for farming and gardening expose humans to such carcinogenic chemicals through the food chain, air, and water supply. There is also evidence of differences in urine organophosphate levels in children with ALL than in controls.

Drugs: Several researchers have linked certain drugs used in chemotherapy for treating other cancers with secondary leukemia [27]. In another study, parental use of diet pills and psychoactive drugs before and during the index pregnancy is associated with increased risks of childhood ALL [28].

2.5.2 Clinical Signs and Symptoms of ALL

ALL patients show different clinical features as a result of marrow failure due to replacement of normal hematopoietic cells by proliferating leukemia blasts. The symptoms are anemia, infection because of neutropenia, gum bleeding and purpura due to thrombocytopia, fatigue because of lymphadenopathy, fever because of hepatomegaly, bone and joint pain due to splenomegaly, and weight loss due to sterna tenderness. Figure 2.4 shows the common symptoms of acute and chronic leukemia that could affect the muscles, skin, lungs, bones, joints and the spleen [29].

2.5.3 Convectional Diagnosis Mechanisms of ALL

Traditionally, leukemia diagnosis is carried out using different medical examinations. One major drawback of these mechanisms is that these medical examination techniques do not detect the disease in its early stages, causing various complications for the patient. Some of these methods are listed below:

A) Patient Clinical History

The main step in making diagnosis is to make a systematic approach to history taking and recording. Competent history taking is part of clinical examination [30]. Such clinical patient information include: present states or symptoms of the patient, past illnesses and family history.

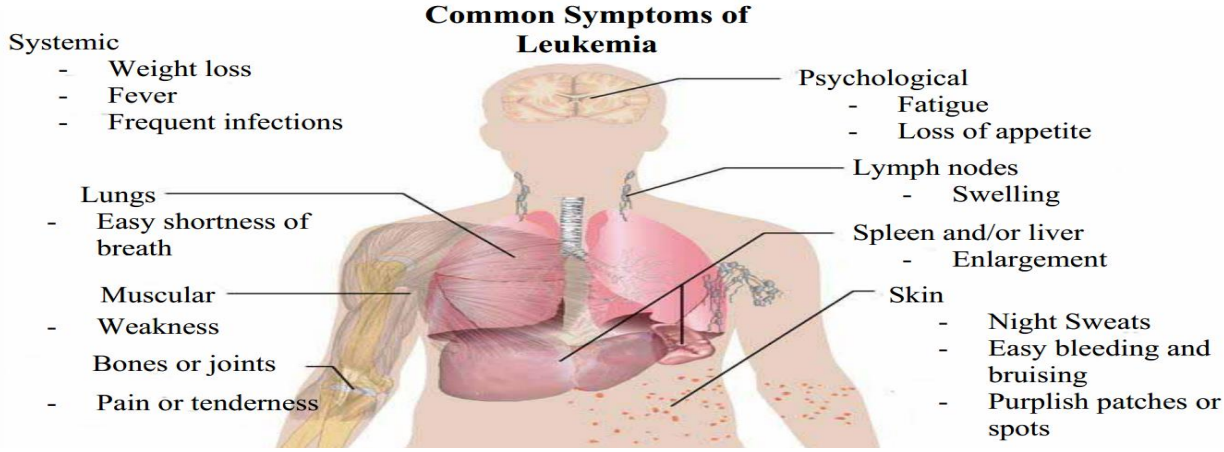


Figure 2.4: Common symptoms of Leukemia.

B) Physical Examination

After reviewing the patient medical history, a physical examination follows in case there is any suspension of leukemia. Under this examination, the clinician look for essential or vital physical signs of leukemia, such as bleeding of the gum, pale skin from anemia, enlarged liver and palpable spleen, and swelling of lymph nodes.

C) Laboratory Examination

The blood laboratory examination evaluates content of the blood component variation under sample test. Leukemia present in someone's blood decreases hemoglobin level and raises WBC count by around 60%-70%. Presence of thrombocytopenia may dictate existence of anemia while peripheral blood smear (PBS) examination reveals around 40-95% blast cells in most of ALL patient [31]. Analysis of cerebrospinal fluid (CSF) also shows the presence of blast cells. The rising of uric acid levels is indicator of high leukemia cell burden of ALL suspects. The modern hematology practices for laboratory diagnosis of ALL depends on blood and bone marrow morphology, immunophenotyping, cytogenetic and molecular analysis.

D) Microscopic Examination

The microscopic examination of blood slides is a standard procedure for ALL diagnosis across the globe. Leukemia blast cells are immature lymphocytes having a completely different morphology than healthy mature lymphocytes and that is the basis of such microscopic diagnosis. The morphological criteria for distinguishing both types of cells are described in Table 2.4. Figure 2.5 presents the general steps that hematopathologists often follow in the lab to diagnose leukemia.

2.6 Limitations of Convectional Diagnosis Mechanisms

Convectional diagnosis is an invasive procedure that provides the physical appearance of the patient, chemistry analysis and display visual images of morphological components of cells under a microscopic study. Microscopic visualization of cellular components even exposes the texture content of cytoplasmic and nucleus regions of the lymphocytes. Hematopathologists have been using laboratory chemistry analysis to count leukemia cells from blood samples and a microscope for the visualization of cell samples. They depend on their clinical expertise while making decisions about the healthiness of the examined blood sample. This includes distinguishing normal mature lymphocytes from leukemic blasts (lymphoblast). Nevertheless, variability in reported manual diagnosis may still occur in all types of cancers including ALL [32]. This could be due to, but not limited to, morphological heterogeneity; noise arising due to improper staining process; intra observer variability, i.e. hematopathologist inability to produce same reading while observing the same samples more than once and inter observer variability, i.e. difference in reading among hematopathologists.

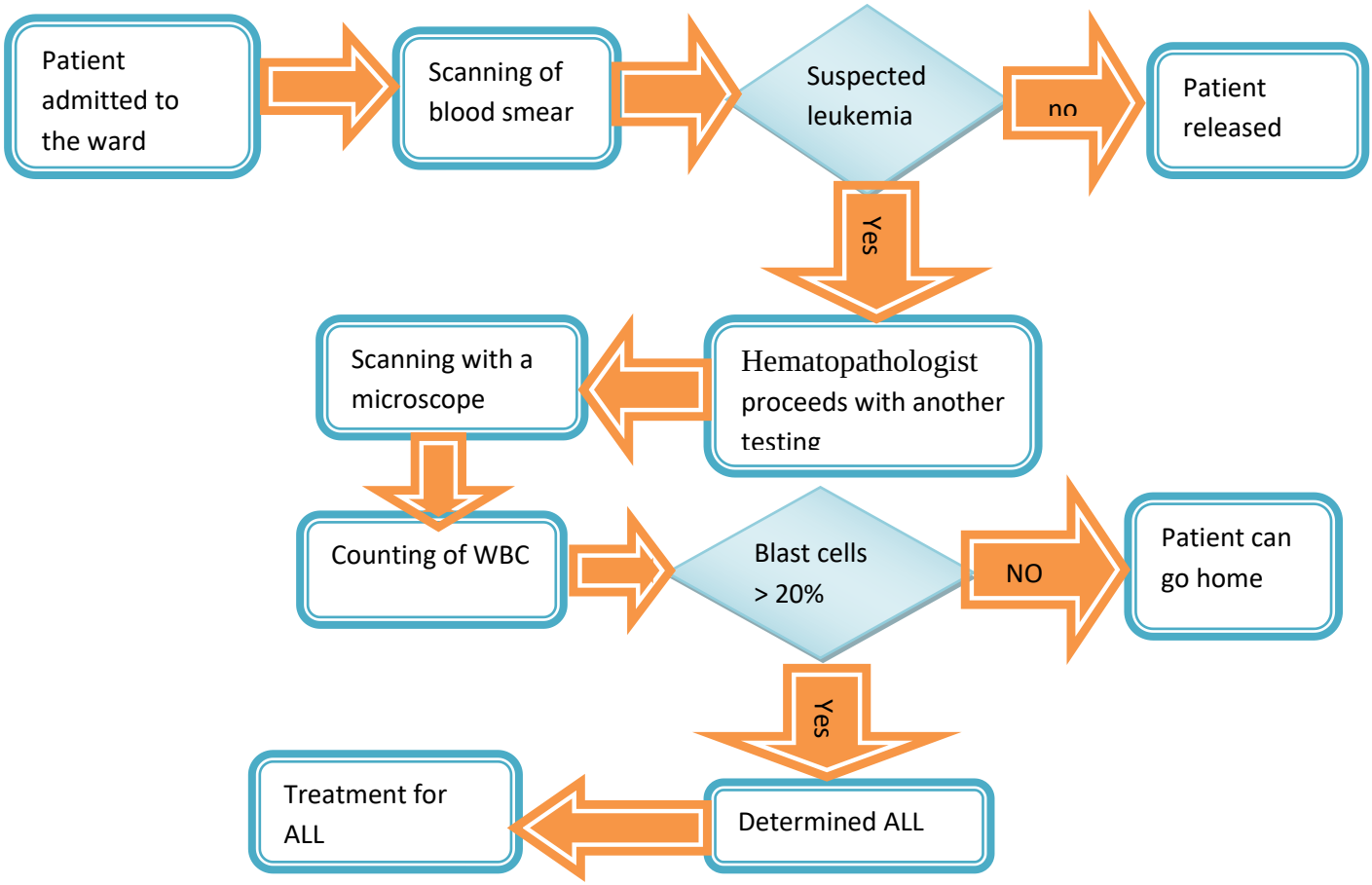


Figure 2.5: General steps that hematopathologists follow in the lab to diagnose leukemia.

Chapter three

3. Hematological Image Processing

3.1 History of Digital Image Processing in Medical Applications

In the late 1960s, digital image processing was facilitated in space applications and in the early 1970s in medical imaging, remote earth resources observations and astronomy. The invention of computed tomography (CT) in the early 1970s is considered a breakthrough in the application of image processing in medical diagnosis. Computational methodologies have improved our ability to analyze and understand images, in medicine and biological sciences. Since then, computers have become an integral part of almost all medical imaging systems including radiography, ultrasound, nuclear medicine and magnetic resonance imaging systems. However, the use of computers and image processing in pathology is quite recent.

With the widespread acceptance of medical imaging as a standard diagnostic tool for various diseases gave an implicit invitation to apply computers and computing for diagnosis of cancer too. Over the last two decades, many image processing based systems have already been designed and successfully used for laboratory diagnosis of various types of cancers. Specifically, computing technology was first applied to microscopic data for the automated screening of gynecological cancer in 1950 [33]. Eventually, with advances in both computing hardware and image processing methodologies, several applications followed: oral cancer [34], ovarian cancer [35], cervical cancer [36], prostate cancer [37], breast cancer [38], colon cancer [39], and follicular lymphoma [40]. In these applications, stained cells or tissue samples are placed under a microscope for scanning and the images of the specific field of view are acquired. Same is true for leukemia detection.

Development of an automated system for cancer diagnosis in the scanned microscopic leukemia images involves four main computational steps: preprocessing, segmentation, feature extraction and detection. These steps are followed by classification of the detected abnormal WBCs from the normal using classification tools.

3.2 Image Formation

An image (assuming grayscale) may be defined as a two-dimensional function $K(x.y)$, where x and y are special coordinates. The amplitude K at any pairs of coordinates (x, y) is called gray level of the image at that point. Alternatively, an image can be seen as a two-dimensional array of numbers [41]. A digital image is composed of a number of elements each of which has a particular location and value. These elements are referred to as picture elements, image elements or pixels. Pixels are widely used to denote the elements of a digital image. Figure 3.1 depicts how a digital image is formed involving a light source, a scene and an imaging system followed by a discretization (sampling and quantization) step.

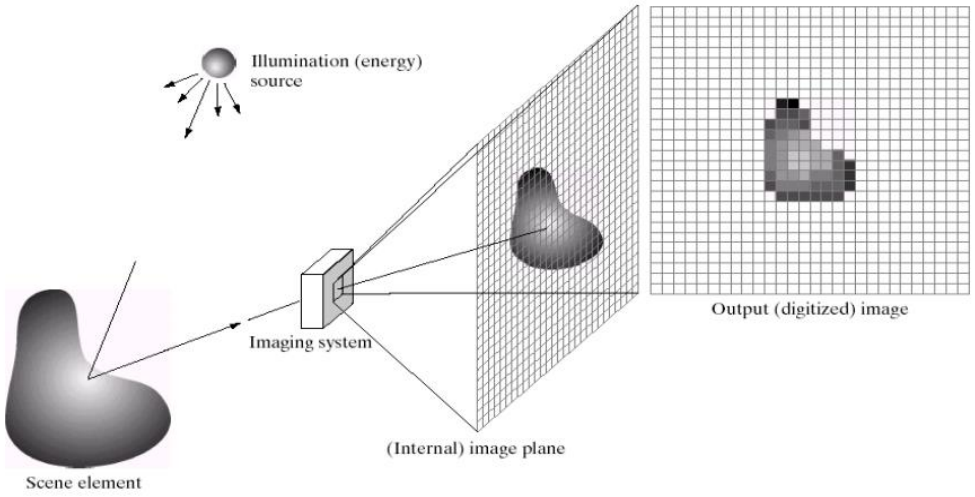


Figure 3.1: Image Formation [41].

The size of the physical area represented by a pixel is called the special resolution of the image. Each pixel has a value and coordinates which reveal its location in the image array. The minimum value a pixel can have is "0" and the maximum depends on how the pixel values are stored in the computer. One way is to store each pixel as a single bit, which means it can take only the values of "0" and "1" or black and white as shown in Figure 3.2. Another common way is to store each pixel as a byte, which consists of 8 bits. In this form, the maximum pixel value is 255.

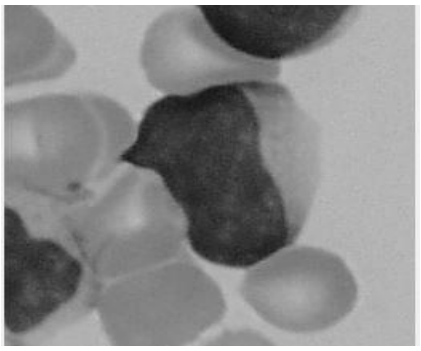


Fig 3.2: Black and white leukemia image.

3.3. Color and Color Space Conversion

Color is a perceptual phenomenon related to the human response to different wavelengths in the visible electromagnetic spectrum. Color is an essential and typical representation for images, and a key element for distinguishing objects. In addition, humans perceive various impressions from color images, such as paintings and photographs. The relationships between color features and human perception are useful in human-computer interaction. Color is one of the fundamental properties of an object and conveys essential information that can be employed in many vision

tasks including but not limited to object recognition, tracking, segmentation, registration and other application. So to properly understanding and use this property we need proper color representation.

A color image is typically composed on three components (bands/channels): the Red, Green and Blue, for example, if it is a RGB type of image. The microscopic leukemia images under consideration are color images with three components. Figure 3.3 presents a typical leukemia image and its three color components. The three components of a color image are like three monochromatic/gray scale images.

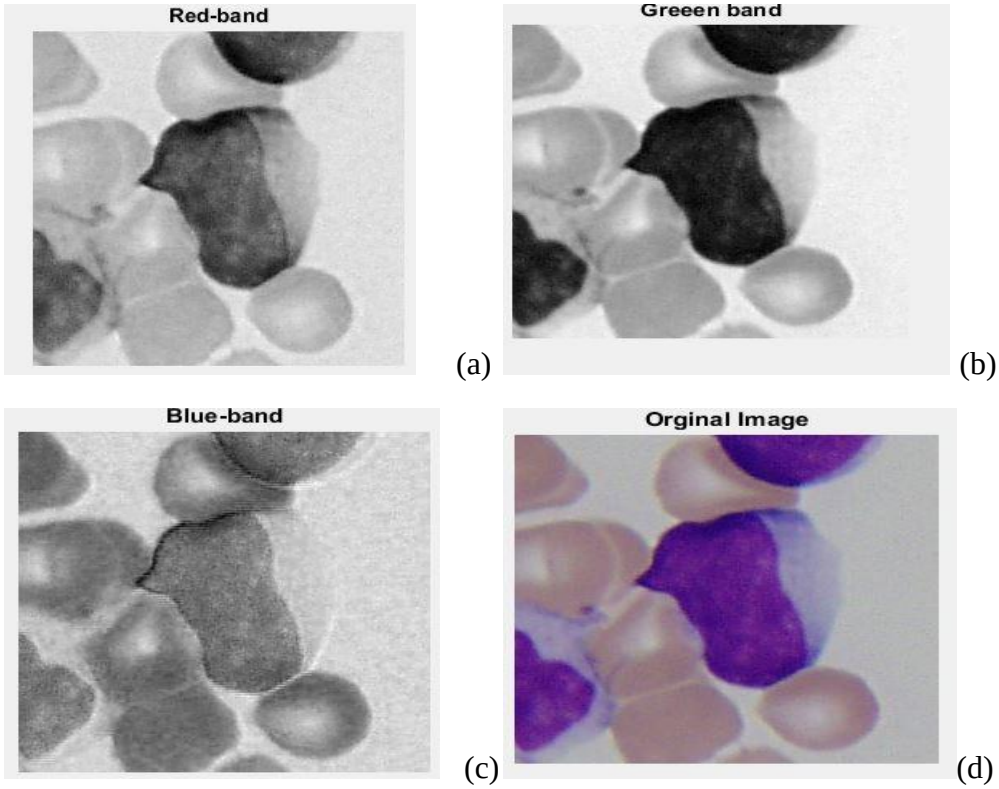


Figure 3.3: Components of RGB leukemia image: (a) Red component, (b) Green component, (c) Blue component and (d) Original RGB image.

A color space is a notation by which we can specify colors, i.e. the human perception of the visible electromagnetic spectrum. Color space is a concept with which color can be specified, created, and visualized [42]. Color spaces are then important as they set the distribution that colors present for different objects, which is fundamental in color classification and segmentation. Color space can be classified as device dependent and device independent. A device independent color space is one where a set of parameters will produce the same color on any equipment. A device dependent color space is a color space where a color depends both on its three coordinates and on the equipment used for display [43].

Manipulation of colors is possible through additive properties (for example, red and green produce yellow) where a wide range of colors is generated from a choice of the three primary colors, i.e. red, green and blue. RGB color space is commonly used in modern displays like televisions, computer monitors and digital cameras [44]. In image processing, this is usually described as a three band/channel system. It means that each color pixel on an image is depicted with three numerical values. Each primary color has a range of values dependent on the bit resolution. Most digital systems store color channels in 8-bit quantities, which allows a range from 0 to 255 to indicate the intensity of colors. This is referred to as 24- bit RGB or true color which assumes 256^3 different colors.

There are other color spaces such as Lab, CMYK, HSV and CIEL L*u*v which might be more important in applications than RGB. The transformation from RGB image to a desired space depends on the specific application [45]. The choice of the color space can be a very important decision because it can dramatically influence the results of the image processing. One major reason to do color space conversion from RGB to HSV, CIEL*u*v or CIEL*a*b targets perceptual uniformity. The HSV color space is an intuitive system, which describes a specific color by its hue (H), saturation (S), and brightness (V). The brightness is also referred as the value or luminosity. This color system is very useful in interactive color selection and manipulation. The CIEL*u*v and CIEL*a*b color spaces are both perceptually uniform systems, which provide easy use of similar matrices for computing colors [46]. Colors like the CMYK are commonly used for printing and not that much in image processing.

3.4. Color Image Segmentation

Image segmentation is one of the early computer vision techniques that involves partitioning an image into a set of homogeneous regions so that the pixels of in each segmented region possess a similar set of properties and attributes [47]. Segmentation procedures mainly depend on various features including color, intensity, textures, shape and other statistical properties. The major intention of image segmentation is to simplify and transform the representation of an image into something that is more meaningful and simpler to analyze. The result of image segmentation is either a set of pixels, regions or objects that cover up the entire image or a set of contours extracted from the image. Those set of segments are used in subsequent image analysis to assisting the depiction, delineation, visualization and classification of regions of interest in any image. For medical images, segmentation is defined as the process of delineating and separating the region of interest from other parts of the image so that it can be viewed individuality in order to achieve important objectives [48]. If the domain of the image is given by Ω, then the segmentation problem is determining the sets whose union is the entire domain Ω. Thus, the sets S_k that make up a segmentation mast satisfy: $U_{k=1}^{M} S_k = \Omega$, where $S_k \cap S_j = \emptyset, \forall k \neq j, k = 1,2, \dots, M,$ each S_k is a connected region, and M is the total number of connected regions. Ideally, segmentation methods find those sets that correspond to distinct anatomical structures or regions of interest in the image.

A great variety of segmentation methods have been proposed in the past decades and the most commonly used approaches are threshold, region growing, classifier based, clustering, Markov random fields, neural network based, edge detection based, model based, physics based and their combinations [49].

Basic Thresholding: Thresholding is an important method of hematological image segmentation and machine vision, which is partitioning an image into homogeneous parts. It is considered an analytical image representation method and plays a very important role in many tasks of pattern recognition, computer vision and image and video retrieval [50]. The main problem that affects this technique is that it only considers the intensity of pixels, ignoring any relationship between them. It may include unimportant pixels values that are not part of the desired region (region of interest) and at the same time miss isolated pixels which are near the boundaries of the region. When thresholding is used, we typically have to play with it, sometimes losing too much of the region and sometimes getting many extraneous background pixels [51].

Clustering: Clustering attempts to segment an image into clusters having pixels with similar characteristics. Data clustering is the method that divides the data elements into clusters so that data elements in the same cluster are more similar than other clusters. In general, clustering techniques can be broadly classified as hard clustering and soft clustering.

The well known clustering algorithms, K–Means, belongs to the first category, hard clustering, where each pixel in an image can belong to exactly one cluster. However, in real situations, there are many probabilities where the clusters are not disjoint and a pixel may have finite belongingness to different clusters. Because of these, soft clustering algorithms have been developed to solve the above limitations and offer a better alternative to hard clustering approaches. In soft clustering, pixels can have partial membership to different classes with a given probability. For example, for microscopic images containing lymphocytes, a particular pixel on the nucleus-cytoplasm boundary has a finite probability of belonging to both the classes, i.e. nucleus and cytoplasm. Some of these soft clustering schemes include Fuzzy C-Means, Rough C-Means, and Shadowed C-Means. One major issue with soft clustering techniques is that the class separation is not well defined and the techniques are mainly efficient in handling overlapping partitions involving spherical patterns and fail when the structure of the input patterns is non-spherical and complex [55].

Watershed Segmentation: A watershed is formed by ‘flooding’ an image from its local minima and forming ’dams’ where waterfronts meet. When the image is fully flooded, all dams together form the watershed of an image. The watershed of edginess’ image (or, in fact, the watershed of the original image) can be used for segmentation. The drawback of watershed based segmentation is over-segmentation due to the frequent presence of multiple markers per region resulting from a poor initialization.

Otsu's Thresholding: The Otsu method is a well known in computer vision and image processing. For a given gray scale image, it finds an optimal threshold for separating an image into two objects, the background and foreground (Otsu, 1979). The Otsu method uses a histogram to represent the foreground and background in the image and the valley points as the threshold and the optimal threshold value is automatically selected.

Related work

The microscopic blood image segmentation is the basis for all automated image based hematological disease detection, recognition and classification systems. Until now, several segmentation methods have been proposed for segmenting WBCs in general. However, several drawbacks are associated with the existing segmentation methods for segmenting WBCs from microscopic blood images, a quick review of which is presented below.

Madhloom et al. [52] proposed automated threshold operations and some image arithmetic operations to localization of the nuclei of WBCs. They select some threshold value to localize the nuclei of the cell. The main problem with this system was the selection of appropriate threshold value for segmentation of the WBCs from the microscopic blood components and the developed system didn't provide sufficient results for the segmentation of WBCs nuclei.

Liao and Deng [53] provided another segmentation method for WBCs from microscopic blood images. They used thresholding to select the WBCs and then used contour identification technique for the WBC cells. The challenging issue was that they assumed that all WBC cells are circular and they assumed the same threshold value for all cells. Some types of WBCs are irregular in shape (Exa. Lymphoblasts) and the idea of using same threshold value didn't work. As a result their system wasn't effective for all WBC cells.

In another study, Halium et al. [54] developed a method for automated blasts counting in leukemic microscopic blood sample images. They applied some thresholding technique by converting the RGB color images into HSV color space. The threshold operation was applied on the S (saturation) component of in the HSV color space to detect WBCs. The major problem with the method was that it didn't include a way for selecting optimum thresholds for better segmentation. The method considered neither feature extraction nor classification.

Mohapatra et al. [56], clustering was applied for WBC segmentation and based on features like shape, color, texture, fractal, Fourier descriptors and contours. The system was trained to recognize leukemia. The results achieved were quite good but a proprietary data set was used to test the scheme that can't be compared with other methods. In another study, Chinwaraphat et al. [57] proposed a method by modifying the Fuzzy C-means clustering technique. The modification has improved the clustering by reducing the uncertainty between the nucleus and cytoplasm selection. The problem with this improved technique was that the researchers had to do some

manual cropping. The improved technique has only been compared to the old technique of Fuzzy C-means.

Umpon [58] proposed a technique for WBC nucleus segmentation using Fuzzy clustering. This method worked good for nucleus segmentation but excludes the cytoplasm, which was the main drawback of the method. Shitong et al. [59] developed a technique merging thresholding and fuzzy segmentation with some mathematical morphology. The method was quite effective in detecting WBCs but not in separating the nucleus and the cytoplasm properly, the major drawback of the technique.

Ghosh et al. [60] proposed a method by using a marker controlled watershed segmentation which segmented the entire WBCs from the background but didn't consider extracting the nucleus and cytoplasm from the background. Dorini et al. [61] implemented watershed transformation techniques for nucleus extraction. They separated the cytoplasm by using the size distribution information assuming the shape of the cytoplasm is round. This information is not true because the cytoplasm is not round.

Gupta el al. [62] developed a system using relevant vector for identifying lymphoblasts. The system identifies three types of lymphoblasts. The system works well for 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 lymphoblasts and had limitations in its effectiveness.

It is noticed that large number of methods are only working on, use gray level analysis of the color microscopic image and use separate analysis of the three channels of the color microscopic images. This leads to a loss of the inter-correlation information embedded within the color bands and the maximum computational cost and time needed when analyzing separate color components.

3.5. Feature Extraction

Feature extraction in image processing involves transforming a set of data into features like textures, color or other geometric entities for use in applications including object classification, pattern recognition and image segmentation. Feature extraction techniques analyze image attributes to extract the most important features from the ROI that are representative of the different classes of the objects.

In image processing, feature extraction involves a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information), then the input data will be transformed into a reduced set of features (also named as feature vector). If the features extracted are carefully chosen, it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input [63]. Thus feature extraction involves simplifying the amount of resources

required to describe a large set of data accurately. When performing analysis on complex data, one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power and may cause a classification algorithm to over fit to training samples and generalize poorly to new samples [64].

Textural Features: Texture is an innate property of virtually all surfaces; the grain of wood, the weave of fabric, the pattern of crops in a field, etc. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Although it is quite easy for human observers to recognize and describe in empirical terms, texture has been extremely refractory to precise definition and analysis by digital computers. Since the textural properties of images appear to carry useful information for discrimination purposes, it is important to develop features for textures.

Texture is an important element to human vision; it provides clues to scene depth and surface orientation. People also tend to relate texture elements of varying size to a plausible 3D surface. If images are represented in grey levels, texture becomes a crucial feature, which provides indications about scenic depth, the special distribution of tonal variations, and surface orientation [65]. In this sense, a texture can be defined as a regional property which is characterized by the spatial distribution of gray levels in a neighborhood. A texture is a visualization of complex patterns composed of spatially organized, repeated sub-patterns, each of which have a characteristic somewhat of uniform appearance. The elements of the patterns are sometimes called textons. The size, shape, color and orientation of the textons can vary over the different regions inside the images. Different types of textures are used for identification and classification of objects. In deterministic texture, the patterns are strictly ordered. In stochastic textures, the spatial distribution of the pattern is random. When the texture patterns have sub patterns within themselves, we call them micro textures.

Texture analysis has been applied extensively in different image processing areas including space sciences, geosciences and optical physic. Different approaches exist to extract textural features from images including statistical, structural, model based and transform based techniques [66]. One of the most commonly used statistical texture analysis techniques is Grey Level Co-occurrence Matrix (GLCM). Meaningful statistics are extracted based on the GLCM of an image as a representation of texture. Since basic texture patterns are governed by periodic occurrences of certain grey levels, co-occurrence of grey levels at predefined relative positions can be a reasonable measure of the presence of texture and periodicity of the patterns. Several textural features such as entropy, energy, contrast, sum mean, variance, cluster shad, cluster prominence, and homogeneity can be extracted from the co-occurrence matrix of gray levels of an image [67]. There also exist structural and model based approaches for texture analysis. Other techniques that have shown great promises in the fields of texture analysis are transform based

approaches. Some of these transforms include the Fourier transform, Cosine transform, Gabor transform and Wavelets.

In medical imaging, textural analysis has proved itself to be are very essential for diagnosis purpose. Visual analysis of textures is often a difficult task, especially in medical images. In general, the texture patterns may be very subtle and their assessment very subjective. Additionally, the basic pattern and repetition frequency of a texture sample could be perceptually invisible. The subject of interest in our case, i.e. color texture analysis, is an important and useful area of study in machine vision and many other applications.

Feature Selection: One of the most common challenges in feature extraction is that large number of features might be extracted which contain irrelevant and redundant features in a specific application. This is where feature selection is primarily performed to select relevant and informative features. Some common feature selection approaches are wrapper methods, Forward Vs Backward Selection, filter approach and others [68]. There are many potential benefits of feature selection including: facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times and defying the curse of dimensionality to improve prediction performance.

3.6. Classification

Classification refers to assigning an image, features of an image or pixels into one of a set of predefined classes. In image processing, the objective of classification is to identify characteristic features, patterns or structures within an image and use these to assign them to particular classes. This is mainly accomplished by using a decision or discriminating function. Image classification is an important and challenging task in various application domains, including biomedical imaging, biometry, video surveillance, vehicle navigation, industrial visual inspection, robot navigation, and remote sensing [69]. Digital image classification techniques can be classified into supervised and unsupervised classification techniques. Unsupervised classification divides unknown image or pixels into distinct classes based on their natural grouping presented in the image. Supervised classification requires the analyst training data to define the classification categories and then classify unknown data [70].

Image classification is one of the basic tasks in image mining and solves the problem of automatic image category annotation. The goal of image classification is to group images into high-level semantic category based on low-level visual features. In image classification, the most important task is to find characteristics of image features for discriminating one category from he other. After defining domain-specific characteristics of images, there are a number of classification techniques that can be used to classify images including decision tree, Naive Bayesian classifier, K-Nearest Neighbor, Multilayer Perceptron, Radial Basis Function Network, Case-Based Reasoning, Rough Sets, Genetic Algorithm and more.

Image classification techniques mainly include image feature extraction and classification methods. The features used to represent medical images can be sorted in two ways: global features and local features [66]. In medical image classification, classifiers are used to divide the feature space into different classes based on feature similarity. Depending on the number of classes, each feature vector is assigned a class label which is a predefined integer value and is based on the classifier output. Each classifier has to be configured such that the application of a set of inputs produces a desired set of outputs. The entire measured data is divided into training and testing data sets. The training data is used for updating the weights and the process of training the classifier is called learning paradigms. The remaining test data are used for validating the classifier performance. In the current thesis work, we propose the use of ANN classifiers for labeling each lymphocyte sub image as normal or malignant sample based upon a set of measured features. Now we briefly describe some of the commonest classification techniques available in the literature.

a) Naive Bayesian Classifier: Naive Bayesian Classifier (NBC) is based on Bayes' theorem and is an important supervised statistical classification technique used in pattern recognition. The working principle of such a classifier is based on Bayes decision theory and the principle of decision is to choose the most probable one [71]. It is designed specifically for classification tasks with features that are independent of one another within each class.

b) K-Nearest Neighbor (KNN): K- Nearest Neighbor (KNN), even though a simple classifier yet yields good classification result. Using KNN classifier each unknown test sample is assigned to a class to which majority of its K-nearest neighbors belong [47].

c) Multilayer Perceptron: Multilayer Perceptron (MLP) is the most popular supervised neural classifier for which many learning paradigms have been developed and are capable of performing nonlinear mapping. In MLP networks, there exists a nonlinear activation function. The hidden layers along with the connected synaptic weights make the MLP network suitable for such nonlinear mappings [72]. Back propagation is generally a supervised method for iteratively calculating the weights and biases of the MLP network

d) Radial Basis Function Network: Radial Basis Function Network (RBFN) has gained considerable attention as an alternate to MLP trained by the back propagation algorithm [73]. The basic functions are embedded in a two layer neural network, where each hidden unit implements a radial activated function. There are no weights connected between the input layer and hidden layer. Finding the appropriate RBFN weights is called network training.

e) Support Vector Machine: Support Vector Machine (SVM) has the capability to distinguish two classes. SVM first uses a nonlinear mapping function for transforming the input data from the observation space to a higher dimensional feature space, and then creates a maximum margin hyper plane to separate the two given classes. Nonlinear mapping functions transform the nonlinear separation problem in the input plane into a linear separation problem in feature space facilitating easier classification in the higher dimensional feature space [74].

3.7. Performance Analysis

Performance evaluation is mandatory in all automated disease recognition systems used to evaluate the performance of a given classifier by comparing it with available ground truth information. The performance test in the current study is conducted to evaluate the ability of the proposed classification scheme for the screening of leukemia in macroscopic blood images. There are different such measures including specificity, sensitivity and overall accuracy. In practice performance metrics are calculated from a confusion matrix. The confusion matrix basically relates the actual (ground truth information) with predicated values.

Chapter Four

4. Proposed Trinion Based WBC Segmentation

4.1 General Overview on the Proposed Method

The proposed technique uses a holistic representation of the color microscopic images in the three (trinion) space and applies trinion based Fourier transforms to extract useful imaging features for the purpose of segmentation and classification of blood microscope images in order to detect ALL. A suitable color space transformation, selection of robust higher order statistical features (Haralick in form), extraction of morphological features, and finally a classification scheme are included in this study. Figure 4.1 shows a block diagram of the proposed method.

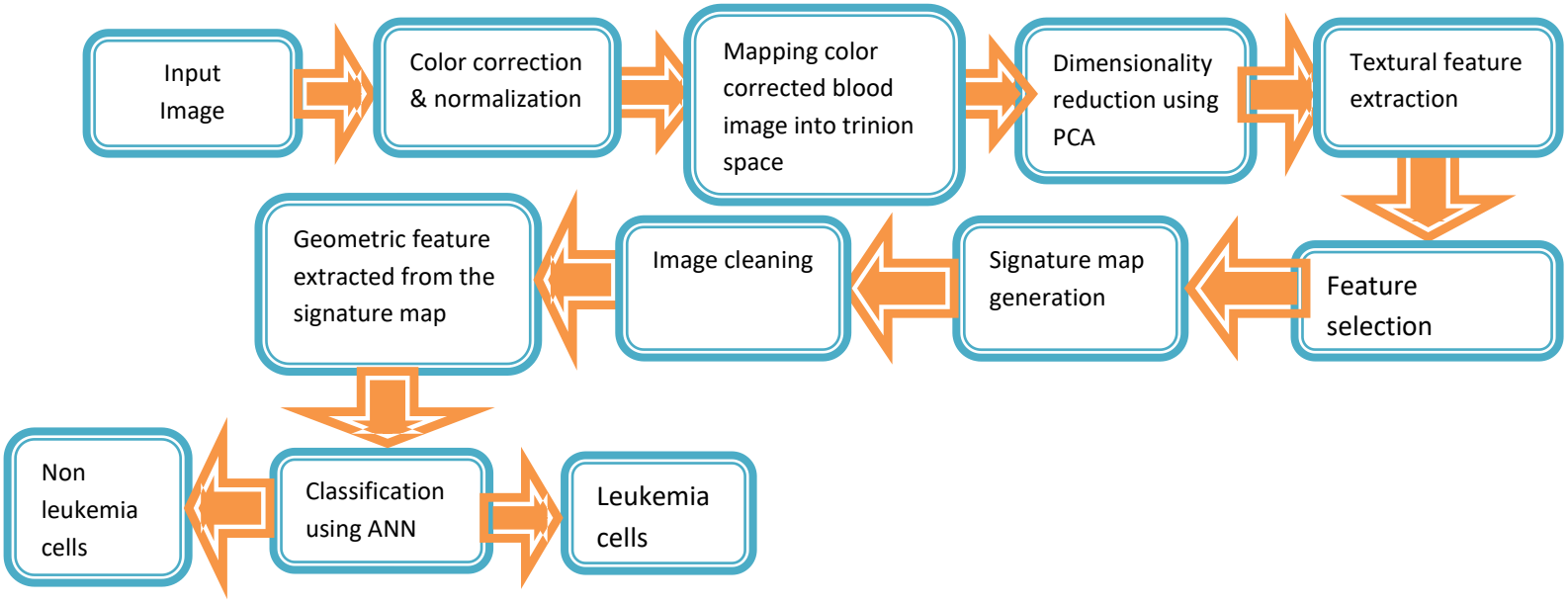


Figure 4.1: Block diagram of the proposed method.

4.2 Image Acquisition

The databases are composed of high quality images accompanied by useful hematological findings. ALL_IDB is a public supervised image dataset of peripheral blood samples of normal individuals and leukemic patients. These samples were collected at a research center for childhood leukemia and hematological diseases, in Monza, Italy. The dataset is used to test and compare cell segmentation algorithms and classification of ALLs.

The images included in the dataset have been captured with an optical laboratory microscope coupled with a Canon Power Shot G5 camera. All images are in JPG format with 24 bit color

depth and 2592 × 1944 resolution. The images are taken with different magnifications of the microscope ranging from 300 to 500. ALL-IDB2 versions of the database are used in the current study. ALL_IDB2 is a collection of cropped areas of interest of normal and blast cells. It contains 260 images with dimension 257 by 257. The data set is public and free available. Figure 4.2a, b: shows an example of the ALL-IDB2 image plotting 2 probable blast cells, and 2 normal WBCs [75].

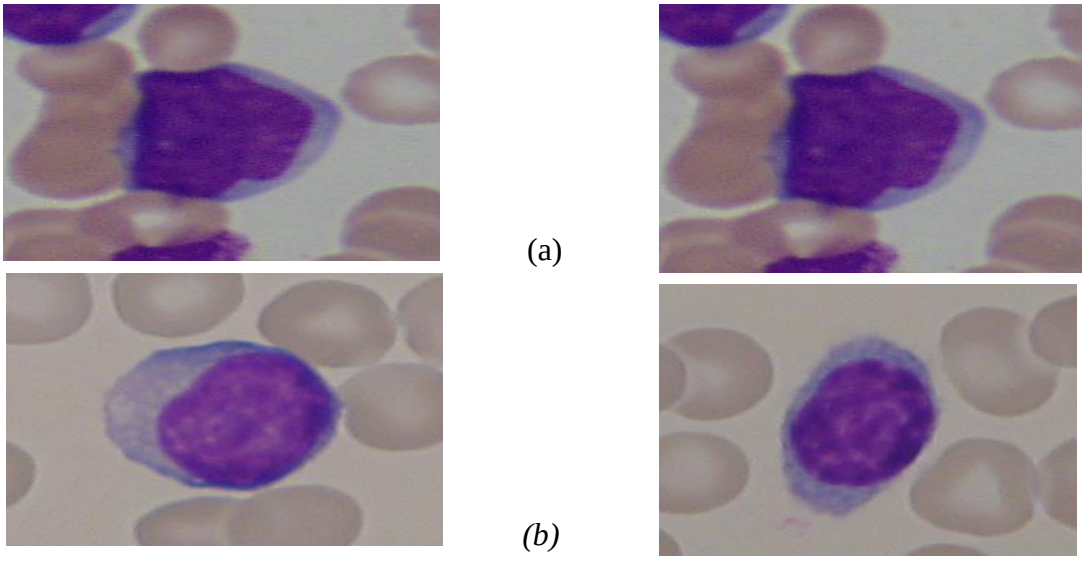


Figure 4.2: (a) Examples of blast cells, (b) Examples of normal WBCs.

4.3 Image Pre-Processing

The microscopic blood images were acquired in RGB color space. Usually the blood cells and image background vary greatly with respect to color and intensity. This is caused by multiple reasons such as camera setting, varying enlightenment, and aging blemish. The intensity variation between the blood cells and image background can be modified by taking the normalized RGB image. Normalized RGB space is formed independently from varying lighting levels. The red, green and blue components of normalized RGB space can be obtained from the three components of RGB space using the following formula:

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \tag{4.1}$$

Colorimetric transformation of the initial color coordinate system is necessary to obtain a color space in which the representation of the color information is appropriate to optimally perform the segmentation process.

The RGB color model is not suitable for stained microscopic blood image segmentation, due to strong correlation among the individual color space planes. Hence, it was necessary to do a color space transformation in order to improve the color variation. In our case, the L*a*b color space

was favored. The L*a*b color space is believed to be suitable for color based image processing as the color dimension is reduced from three to two and the color channels are uncorrelated. This colorspace consists of a luminosity layer L* and chromaticity layers a* and b* and the color information is embedded in these two layers.

4.4 TFT Based WBC Segmentation Using Textural Features

There are many techniques applied for microscopic blood image segmentation to detect WBCs. Some of the methods include automatic thresholding, Fuzzy-c means, K-means clustering, watershed transform, edge based techniques and so on. These techniques are mainly applied on gray level images but also used for colors. The traditional approach followed in applying these methods on color images involves separation of the color components which misses the intrinsic inter correlation information embedded among the color bands/components. The other drawback is to do with increased computational cost when analyzing separate color components. This calls for the development of a more holistic approach.

One novel approach that was proposed in the literature for more effective analysis of color images makes use of vectorial representation of the color pixels. In this regard, there are two representations that are used namely quaternions (in four space) and trinions (in three space). The quaternion and trinion based integral transforms are used to effectively and efficiently analyze color images by treating the color pixels as one entity using vectors [76].

A quaternion is defined with one real and three imaginary components and has been used to represent three bands of color images by mapping to the three imaginary components of a quaternion setting the real component to zero. However, most color images, including microscopic blood images, have three components and hence the extra dimension in quaternions might be redundant. Such redundancies could be handled using trinions and the approach has been adopted in this thesis study to develop the WBC segmentation scheme to automatically detect leukemic cells [77].

Trinions are defined based on principles of Clifford algebra and come with one real and two imaginary components. A trinion number is defined as:

$$t = a + ib + jc \tag{4.2}$$

where a, b, c are real numbers and i and j are operators that fulfill the condition $i^2 = j, ij = ji = -1$ and $j^2 = -i$. The three base elements {1, i, j} of trinions form a commutative group where 1 is the multiplicative identity element. Distinct from quaternions, trinions with the above structure form a commutative ring. Trinion are associative as well as distributive with respect to addition and multiplication [77]. Any trinion number *t* can be expressed as the sum of a real part and a vector part as:

$$t = S(t) + V(t) \tag{4.3}$$

where $S(t) = a$ is the real part and $V(t) = ib + jc$ is the vector part. It can also be expressed in sin and cosine form as:

$$t = |t|(\cos\varphi + \mu \sin\varphi) \quad (4.4)$$

where $[t] = \sqrt{a^2 + b^2 + c^2}$, $\mu = \frac{V(t)}{[V(t)]}$ and $\varphi = \left(\frac{|V(t)|}{S(t)}\right)$, 0 < φ < π are the amplitude (modulus), the Eigen axis and Eigen angle (phase), respectively. When $|t| = 1$, it is a unit trinion, when a=0 it is a pure trinion. Other properties of trinion are explain in the literature [77].

Two general working definitions for the trinion Fourier transform (TFT) have been suggested in [78]. These are type-I and type-II. The TFT type-I and its inverse (ITFT) are given by:

$$T(u\ v) = \iint_{-\infty}^{\infty} h(x\ y)(\cos\big(2\pi(ux + vy)\big) - \mu_1 \sin(2\pi(ux + vy)))\, dx\, dy \quad (4.5)$$

$$h(x\ y) = \iint_{-\infty}^{\infty} T(u\ v)(\cos\big(2\pi(ux + vy)\big) + \mu_2 \sin(2\pi(ux\ vy)))\, du\, dv \quad (4.6)$$

where h(x y) is generally a trinion valued image function, $\mu_1$is a unit pure trinion, and μ_2 is a trinion so that the product $\mu_1\mu_2 = -1$. The selection of μ_1 and μ_2 is arbitrary and one could use by $\mu_1 = \frac{(i-j)}{\sqrt{2}}$, and $\mu_2 = \frac{(1-i+j)}{\sqrt{2}}$, as describe in the literature [77].

The discrete form of type-I TFT has been effectively applied in different image processing applications including brain tissue classification of multi-parametric magnetic resonance images [78], robust identification of lesions on retinal color image processing [79], three-dimensional wind profile prediction [80], and edge detection in medical images with multiple components [81]. Readers may consult the literature to know more on type-I TFT as well as type-II TFT.

The type-I TFT has been adopted in this thesis as a basis to effectively classify normal and leukemic WBCs from the microscopic images. The mapping of the color images into the trinion space was done following color space transformation from the RGB to L*a*b. The discrete form of type-I TFT was applied locally over a spatially translating window of size 3x3 and feature extraction was performed to extract surrogate markers for normal and abnormal WBCs.

4.4.1 Data Dimensionality Reduction Using PCA

Dimensionality reduction is the introduction of new features space where the original features are represented. The new feature space is of lower dimension than that of the original space. It involves feature selection and feature extraction. This makes analyzing data much easier and faster for machine learning algorithms without extraneous variables to process. There are different types of dimensionality reduction techniques including generalized discriminant analysis (GDA), linear discriminant analysis (LDA), principal component analysis (PCA) and non-negative matrix factorization (NMF). Among these the most classic technique is PCA and that was adopted in this thesis. PCA was applied locally over each 3x3 trinion transformed

microscopic blood image. Then the values of the resulting 3x3 matrix were normalized between 0 and 1 and these are the probability density functions (PDF) used to compute textural features for use in uniquely quantifying the normal and leukemic WBCs.

4.4.2 Textural Feature Extraction and Generation of Signature Map

A feature vector is an n-dimensional vector of numerical features that represents some objects. Many algorithms in machine learning require a numerical representation of objects, since such representations facilitate processing and statistical analysis. When representing images, the feature values might correspond to the pixels of an image. The n-dimensional vectors may be first, second or higher order numerical features that represent the image or regions inside the image.

The texture relates mostly to a specific, spatially repetitive structure of surfaces formed by repeating a particular element or several elements in different relative spatial positions. Generally, the repetition involves local variations of scale, orientation, or other geometric and optical features of the elements. Hence the image texture gives information about the spatial arrangement of color or intensities in the image or selected regions of the image. Many statistical texture features are based on co-occurrence matrices representing second-order statistics of grey levels in pairs of pixels in an image. In our case, the statistical values of the second and higher order 'Haralick' texture features were extracted from TFT transformed image matrix. These features were 'Haralick' only in forms and the way they are computed was completely different from the way the traditional Haralick features are computed. Once PCA is applied on the trinion transformed image locally over a translating window of size 3x3, the features were computed over each window and the feature value was assigned to the central pixel. This procedure is repeated across all pixel positions that are included in the selected region of interest. Finally, texture/signature maps are generated as color.

About ten Haralick texture feature were computed using formulas described in Table 4.1 and they were tested for their efficacy in detecting WBCs from the microscopic blood images.

Table 4.1: Formulas used to compute Haralick textural features [82].

No	Features	Formula
1	Energy	$\sum_{u=1}^{3}\sum_{v=1}^{3}P(u,v)^2$
2	Sum mean	$\frac{1}{2}\sum_{u=1}^{3}\sum_{v=1}^{3}(u\big(P(u,v)\big)+v\big(P(u,v)\big))$
3	Square mean	$\frac{1}{2}\sum_{u=1}^{3}\sum_{v=1}^{3}(u(P(u,v)^2+v(P(u,v)^2))$

4	Variance	$\frac{1}{2}\sum_{u=1}^{3}\sum_{u=1}^{3}((u-\mu)^2\,P(u,v)+(v-\mu)^2P(u,v))$
5	Entropy	$-\sum_{u=1}^{3}\sum_{v=1}^{3}P(u,v)\log(P(u,v))$
6	Contrast	$\sum_{u=1}^{3}\sum_{v=1}^{3}(u-v)^2\log(P(u,v))$
7	Homogeneity	$\sum_{u=1}^{3}\sum_{v=1}^{3}\frac{P(u,v)}{1+(u-v)^2}$
8	Cluster shade	$\sum_{u=1}^{3}\sum_{v=1}^{3}\left(u+v-\mu_x-\mu_y\right)^3P(u,v)$
9	Cluster prominence	$\sum_{u=1}^{3}\sum_{v=1}^{3}\left(u+v-\mu_x-\mu_y\right)^4P(u,v)$
10	Correlation	$\sum_{u=1}^{3}\sum_{v=1}^{3}\frac{(u-\mu_x)\left(v-\mu_y\right)P(u,v)}{\sigma_x\sigma_y}$

where $P(u,v)$ is the normalized spectral value obtained after application of PCA on the TFT transformed image matrix, $\mu=\frac{1}{9}\sum_{u=1}^{3}\sum_{v=1}^{3}P(u,v)$ is mean of the matrix, $\mu_x=\sum_{u=1}^{3}u\sum_{v=1}^{3}P(u,v)$ is the sum of raw mean, $\mu_y=\sum_{v=1}^{3}v\sum_{u=1}^{3}P(u,v)$ is the sum of column mean, ${\sigma_x}^2=\sum_{u=1}^{3}(u-\mu_x)^2\sum_{v=1}^{3}P(u,v)$ is the sum of raw variance and ${\sigma_y}^2=\sum_{v=1}^{3}(v-\mu_y)^2\sum_{u=1}^{3}P(u,v)$ is the sum of column variance.

The final signature map is generated after computing these features. The best signature map allows differentiating the WBCs from the other blood components (RBC, platelet, and other objects including the image background). As mentioned in the literature review part of this thesis, only the WBCs are relevant to detect ALL. The performance of signature maps generated using the different texture features was compared and assessed accordingly through qualitative comparison of the signature maps against the available ground truth. As presented in the results section, the best signature maps were obtained from the cluster prominence textural feature which allowed a more accurate segmentation of the WBCs compared to the other nine features.

4.5 Image Cleaning on the Generated Signature Maps

The main objects of interest in the microscopic images that are relevant in detecting leukemia are the WBCs. In some instances the generated signature maps may create small holes inside the detected WBCs. Also some of the WBCs detected might appear on the edges of the microscopic images (partial WBCs) under consideration which may make further processing challenging.

During the image cleaning process, the partial WBCs will be removed. In order to perform the image cleaning, different geometric features may be used. In this thesis, a size metrics is used to do cleaning by computing the solidity feature on the generated signature maps. Solidity is computed as the ratio of area to the convex hull given by:

$$Solidity = \frac{Area}{Convex\ Area} \tag{4.7}$$

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. A threshold value of the solidity measures is used for identifying the range of small areas of the WBCs that should be removed from the signature maps.

4.6 Geometric Feature Extraction from the Cleaned Signature Map

The blood cells are of different size and shapes. Shape based descriptors are very useful to differentiate normal WBCs from the abnormal ones and hence to detect ALLs. In order to extract the geometric features from the generated signature maps, first the signature maps are converted into gray levels. After that, thresholding is applied to convert the resulting gray level images into binary form. This is followed by the application of morphological operations including area opening, closing, erosion and dilation. Here, area opening is performed to remove connected components while dilation adds pixels to boundary of objects. Erosion is used to remove pixels on object boundaries. Hole filling operation is performed to detect perfect WBCs. After the hole filling operation, edge detection was applied using the classical approaches. In this thesis work, Canny was applied to detect the edges as it offers better performance than the other classical approaches including Sobel, Prewitt and the like. The edge detection is followed by computation of different shape based features like major axis length, minor axis length, area, perimeter, radius, solidity and Centroid. Figure 4.6 presents the general algorithmic steps followed for morphological feature extraction.

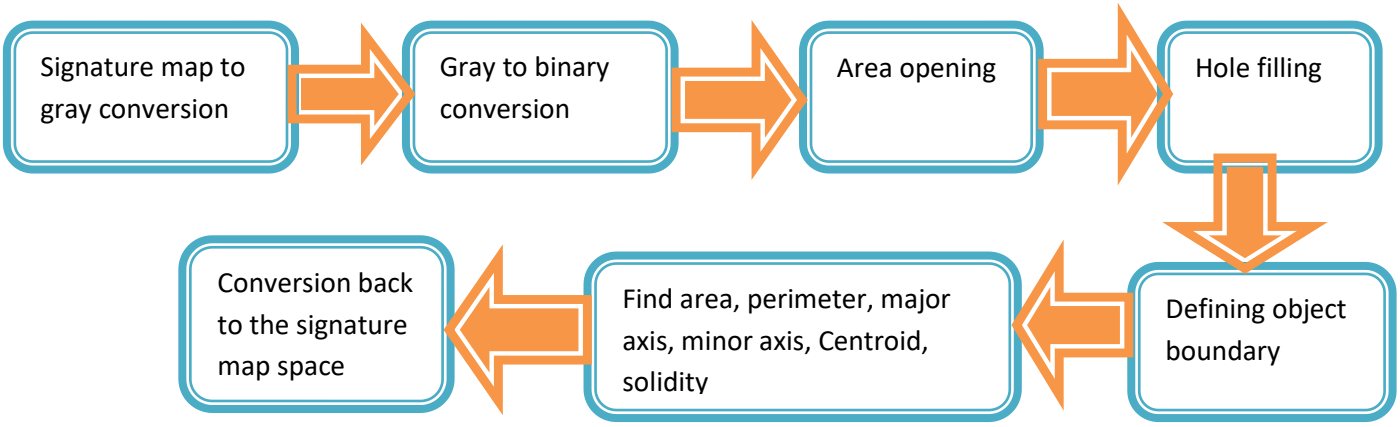


Figure 4.3: The general algorithmic steps followed for geometric feature extraction.

4.7 Classification using Artificial Neural Network (ANN)

Classification is the last step in the experimental methodology presented in this thesis. Over the past decades, there has been growing interest in Artificial Neural Network (ANN), mostly in the area of feed forward networks for pattern recognition applications. Neural networks were invented in the early 1940s by McCulloch and Pitts. ANNs are innovative models of human cognitive function. The artificial equivalents of biological neurons are the nodes or units of neural networks. The human brain consists of an estimated 100 billion nerve cells or neurons as shown Fig. 4.4 [83]. The term "network" is used to refer to any system of artificial neurons. A neuron is a cell in the brain whose main activity is the collection, processing and dissemination of electrical signals. The brain's information-processing function is thought to emerge from networks of such neurons. Neurons communicate through electrical signals that are short-lived impulses or "spikes" in the voltage of the cell wall or membrane. The interneuron connections are mediated by electrochemical junctions called synapses that are located on branches of the cell referred to as dendrites.

Each neuron is typically connected to many thousands of other neurons and continually receives a multitude of incoming signals, which reach the cell body. Signals are integrated or summed together and if the resulting signal exceeds some threshold, it causes the neurons to generate a voltage signal in response. Then it transmits this signal to other neurons through a branching fiber known as the axon. It is this structure and method of processing that neural networks mimic. Each synapses or connection has a weight, hence each input is multiplied by its weight before being sent to the equivalent of the cell body.

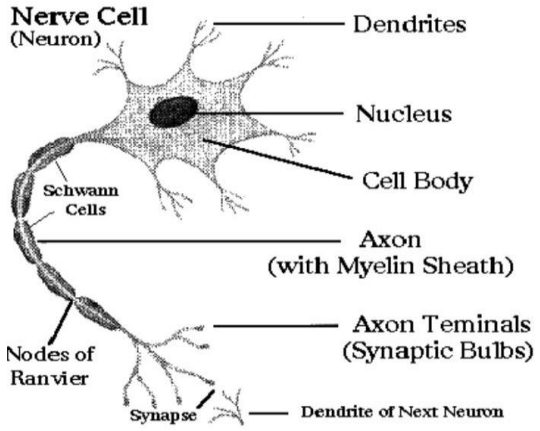


Figure 4.4: Structure of a human neuron [83].

The weighted signals are summed together by basic arithmetic addition to cause node activation. This may range from something as basic as a signal node to a large set of nodes, each one connected to every other node in the net. They are nodes arranged in a layered structure in which each signal from input posses to two nodes before reaching an output. In real neurons, the synaptic strength, under certain circumstances, is modified so that the behaviors of each neuron can change or adapt to its unique stimulus inputs. In artificial neurons, the equivalent of this is

the connection weight value. The “knowledge” of a network is supposed to be stored in its weights [84].

A neural network structure consists of nodes that represent a processing unit with relationships between the units as indicated by arcs shown in Fig. 4.5. There are a number of processing units called hidden units. Each hidden units must be used internally to be connected to the input and output units, using weighted connections and the network must be trained [85]. Neural networks are an attempt to build machines that operate in the same way as the human brain.

Neural network structures belong to two main categories: feed-forward networks and recurrent networks. A feed-forward network represents a function of its current input and weights. A recurrent network feeds its output back into its own input. That means the activation levels of the network build a dynamical system that may influence a steady state. Moreover, the response of the network for a given input depends on its initial state and previous inputs. Feed-forward networks are arranged as layered neural network (or perceptron network) with all the inputs connected directly to the outputs. Each output unit is independent of the others and each weight affects only one of the outputs [86]. The function of a neural network is to produce an output pattern when presented with a suitable input pattern [87]. The basic architecture of a neural network has been depicted in Fig. 4.5.

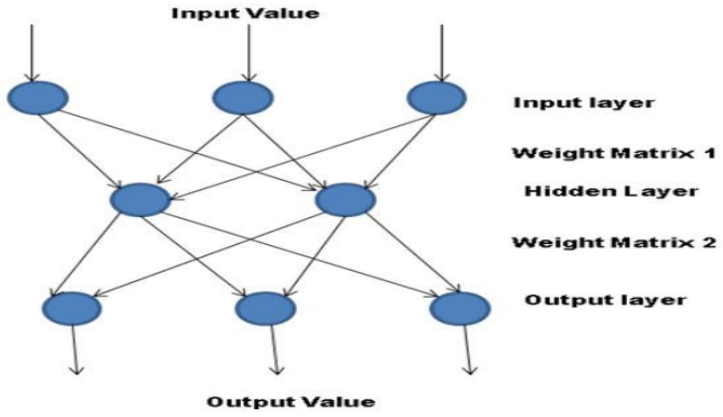


Figure 4.5: The basic architecture of a neural network.

Neural networks are used to solve problems in pattern classification, speech recognition, textual character recognition and fields of human knowledge such as medical diagnosis, geological survey of oil and financial market forecasts [84]. Pattern classification is the process of sorting patterns into one group or another in a large number of cases. In this thesis, ANN classifier is used based on morphological feature data as input and the morphological data are taken from both patient and normal blood samples.

4.8 Classification Performance Measurements

Generally speaking, classification problems are evaluated using a matrix known as the confusion matrix. The confusion matrix contains the numbers of correctly and incorrectly classified samples for each class. Table 4.2 shows confusion matrix containing two classes (positive and negative), which is suitable for binary classification problem (Costa et al. 2007) [90].

Table 4.2: Confusion matrix for two classes.

True class	Predicted class	
	Positive	Negative
Positive	True positive	False negative
Negative	False positive	True negative

In general, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are computed for all outputs in the "classifier testing set" through testing the classifier. FP is the proportion of non-targeting output classified incorrectly as targeting outputs. On the other hand, TP is the proportion of targeting output classified correctly as targeting outputs. TN is the proportion of non targeting output classified correctly as non-targeting output while FN is the proportional of targeting output classified incorrectly as non-targeting outputs. Four matrices are commonly used to evaluate the performance of the targeting output classifier. The first metric is called sensitivity which describes the rate of the TP. It is the probability in which ALL elements are correctly classified as ALL. The sensitivity is defined as:

$$Sensetivity = \frac{TP}{TP+FN} \tag{4.8}$$

The second metric is called specificity; this refers to the rate of TN. It is the probability in which non ALL elements are correctly classified as non ALL. Specificity is defined as:

$$Specificity = \frac{TN}{TN+FP} \tag{4.9}$$

The specificity and sensitivity are used when the performance of both classes are of concern and expected to be high concurrently. The last two metrics are precision and accuracy. Precision shows how consistent the results can be reproduced, while accuracy reflects the overall correctness of the classifier. Precision and accuracy are defined as:

$$Precision = \frac{TP}{TP+FP} \tag{4.10}$$

$$Accuracy = \frac{TP+TN}{TN+FN+FP+TN} \tag{4.11}$$

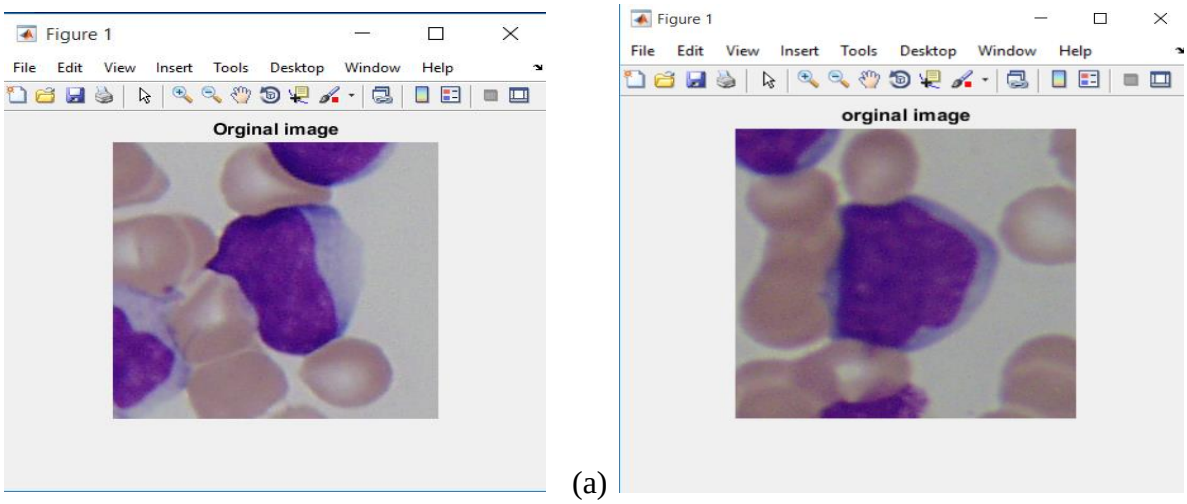
Another very useful evaluation measure used for making a decision about the optimum model is the Receiver Operating Characteristics (ROC), which relates sensitivity and specificity. The area under the curve (AUC) is often taken as a comprehensive scalar indicator of the model performance. AUC value of 0.5 suggests a poorly performing model with random sample assignment, while an AUC value of 1 indicates a model of maximum efficiency.

Chapter 5

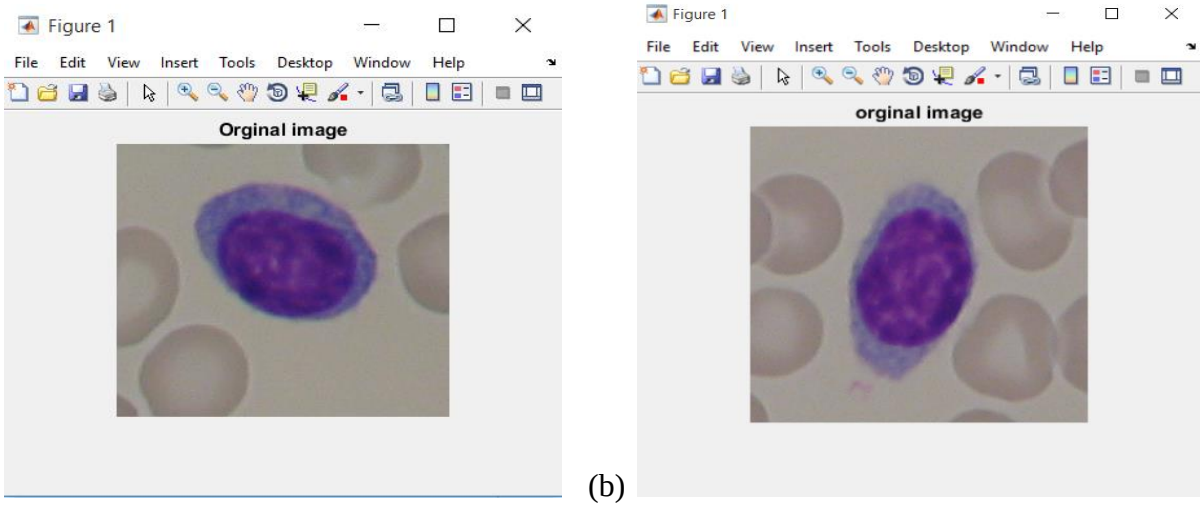
5. Results and Discussion

5.1 Detection of WBCs from Microscopic Blood Images

Before generating signature maps, the original microscopic blood images were subjected to pre-processing steps in order to reduce the effects of illumination, image corruption due to staining irregularities and other artifacts. First, the RGB microscopic images are normalized to reduce possible intensity variations in the images. Then conversion from RGB to L*a*b color space was applied. Figure 5.1 presents typical healthy and unhealthy microscopic blood images in the original RGB and L*a*b color spaces.



(a)



(b)

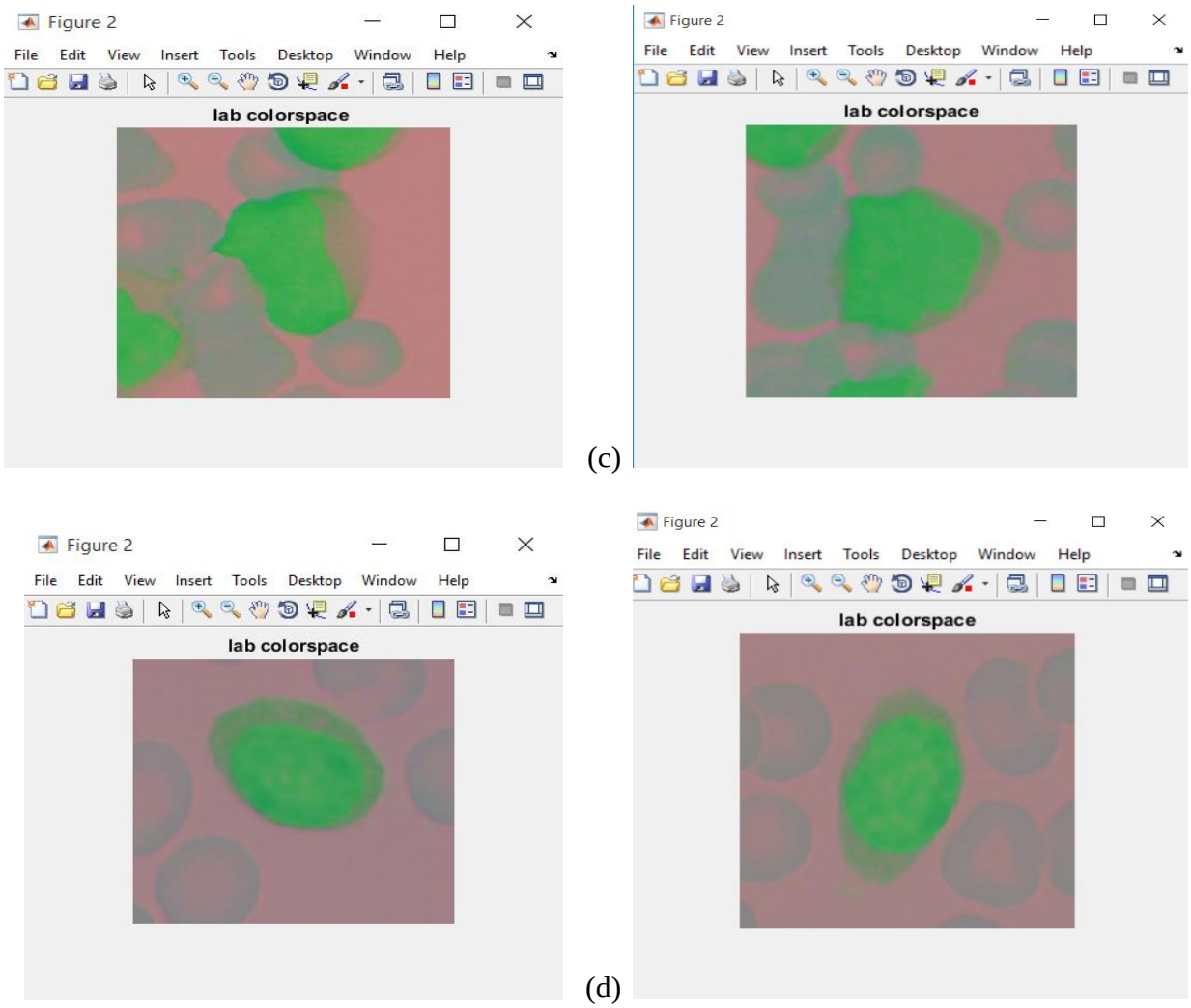


*Figure 5.1: Unhealthy microscopic blood image (a) and healthy microscopic blood image in the RGB space; Unhealthy image (c) and healthy microscopic image (d) in L*a*b color space.*

5.2 Signature Map Generation

The signature map was generated as a color using trinion valued statistical features. Signature maps generated using cluster prominence computed in L*a*b color space showed improved results over the other features in terms of accurately detecting the WBCs. Figure 5.2 presents signature maps generated on both health as well as abnormal (taken from ALL patients) microscopic blood images. In all cases the WBCs were correctly identified by the proposed scheme. The test images considered in the current study were composed of both good quality images as well as other images which are significantly affected by poor illumination as well as noise. Qualitative analysis was done to investigate the potential and limitations of the algorithm by observing the resulting signature maps of all test images included in the sample set. The experimental results have shown that the signature maps were found very helpful for visual

enhancement of pathological features on the WBCs. As demonstrated in Fig. 5.2, the WBCs appear in color on the generated signature maps distinct from the background that appeared dark and all other blood components have been removed on the signature maps. The generated signature maps robustly detected the WBCs in a unique color. Generally the tolerance of the algorithm for change the background color and removal of other blood component was found very satisfactory.

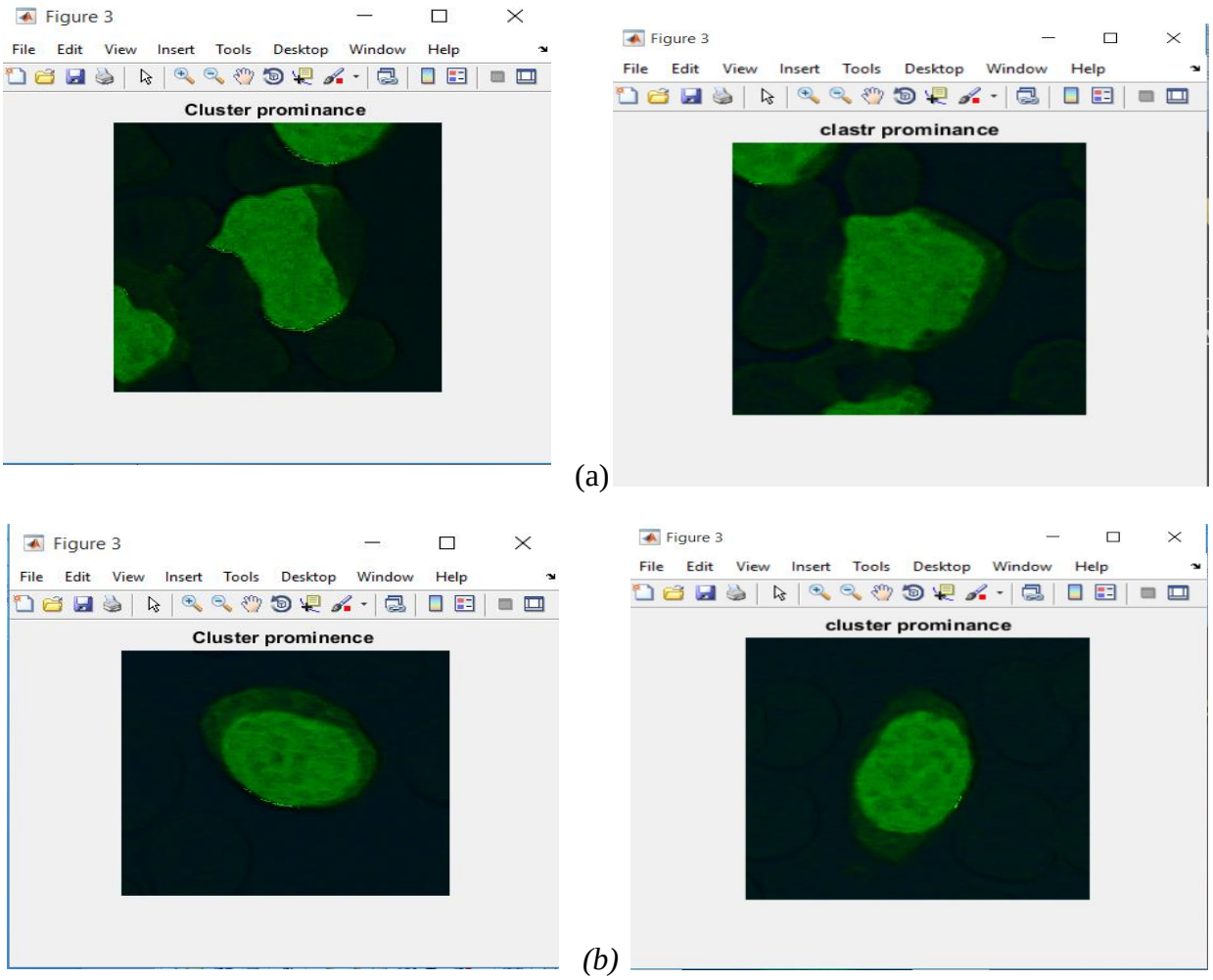


Figure 5.2: Signature map generated based on cluster prominence feature on unhealthy image sample (a), signature map generated based on cluster prominence feature on healthy image (b).

5.3 Image Cleaning on the Signature Maps

The threshold value for solidity which is used for identifying the range of partial WBCs and small areas can be obtained from the signature map. After having so many experiments, it was found that signature map values less than 0.8 are able to efficiently suppress WBCs that exist at the edge of the signature maps. The Appendix presents typical results obtained after the cleaning operation is applied on the signature maps. The components having values below the threshold

are the components which are on the edge of the signature map which need to be discarded. And then the geometric features are extracted from the remaining WBCs which are important to identify the healthy and unhealthy WBCs to detect the ALL. This was done through further morphological analysis. Figure 5.3 presents the general cleaning procedures applied on the signature maps that include suppressing partial WBCs, morphological opening, filling, edge detection as well as the dilation operator.

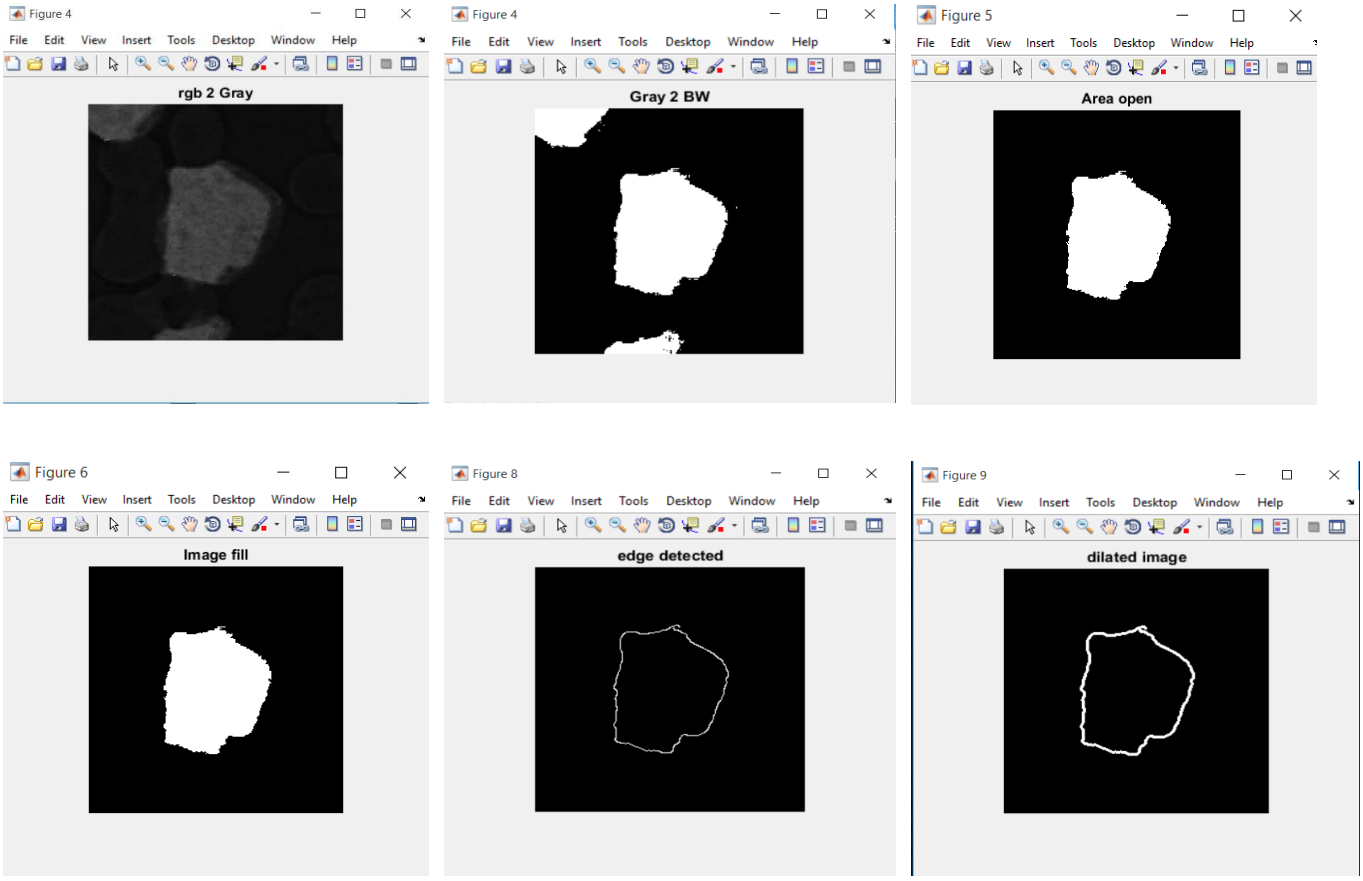


Figure 5.3: General steps of cleaning the signature maps.

Figure 5.4 presents the final images obtained after the morphological cleaning operation is applied for both healthy as well as infected microscopic blood samples.

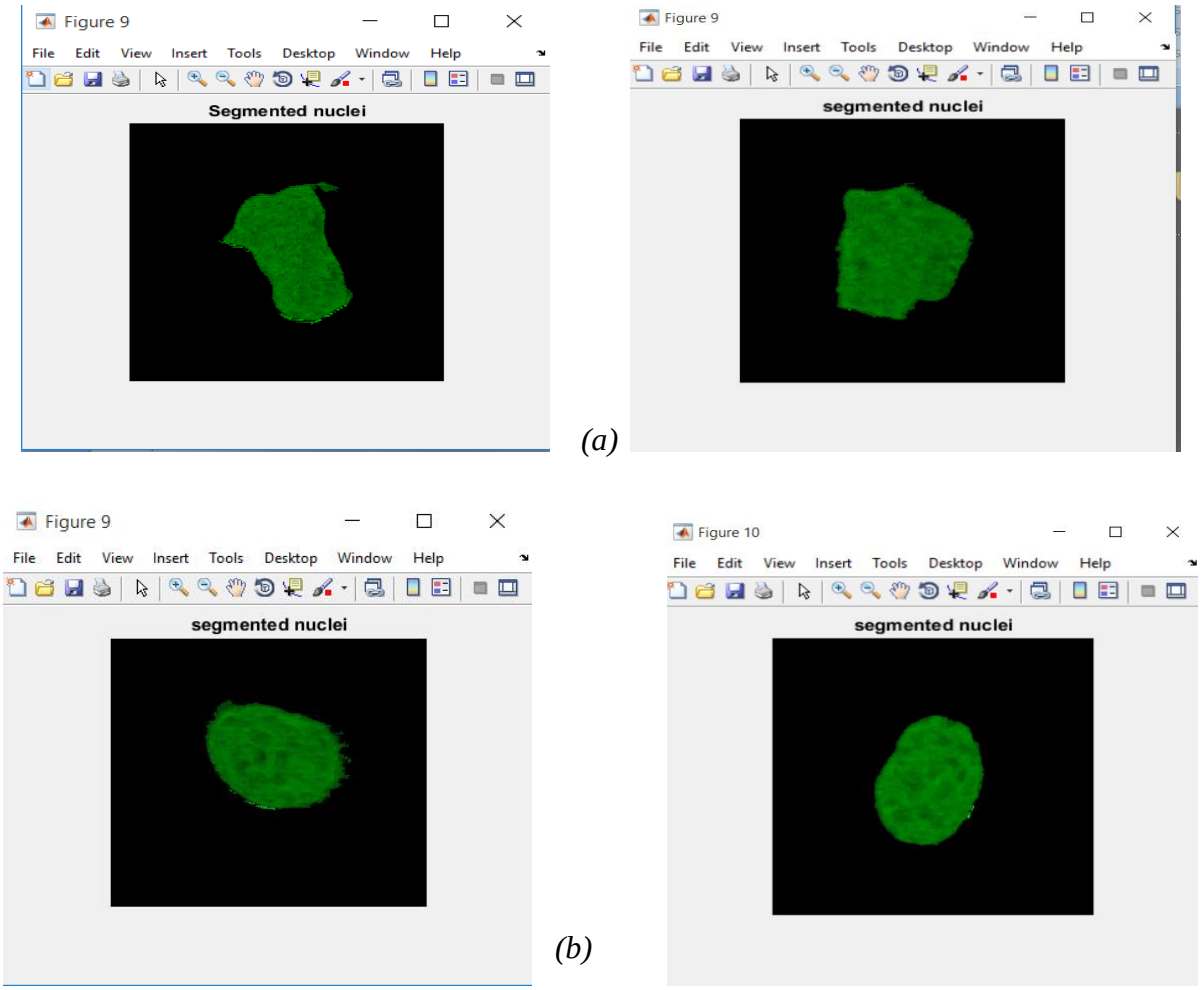


Figure 5.4: The final signature map of the infected WBCs (a) and healthy WBCs (b) after the morphological cleaning operation.

5.4 Extraction of Geometric Features

WBCs contain important information associated with their geometrical appearances. For example, the shapes of different blasts are either round, oval or kidney-shaped. In the current work, the following morphological features of the WBCs that appear on the final signature maps are considered:

- Area: the number of pixels on the interior of the cells defined separately for the normal and infected WBCs. The nuclei of the WBCs are considered in both cases.
- Perimeter: distance between the pair of pixels around the boundary.
- Solidity: the ratio of the area over convex area.
- Major axis length: length (in pixels) of the major axis of the cells that has the same normalized second central moments as the region.
- Minor axis length: length (in pixels) of the minor axis of the cells that has the same normalized second central moments as the region.

- Centroid: 1-by P-vector, the center of mass of the region. The first element of the centroid is its horizontal coordinate and the second element of the centroid is its vertical coordinate.

The value of those geometrical features taken from healthy and infected WBCs nuclei is presented in Table 5.1 and Table 5.2. From the metrics computed in the tables, it is clear that both *Area* and *Perimeter* parameters were able completely differentiate healthy and ALL infected WBCs. The other two features, *Major Axis Length* and *Minor Axis Length* also performed quite well while *Solidity* was inferior in the classification task.

Table 5.1: Geometrical features extracted from selected healthy segmented WBCs.

No	Area	Major axis length	Minor axis length	Solidity	Perimeter
Cell-1	6192	90.98	87.148	0.958	318.41
Cell-2	5377	86.82	79.40	0.959	312.80
Cell-3	6520	96.31	87.62	0.931	373.48
Cell-4	7821	100.82	99.10	0.97	337.11
Cell-5	7763	121.88	81.37	0.973	343.41
Cell-6	5377	86.82	79.40	0.959	312.80
Cell-7	6520	96.31	87.62	0.931	373.48
Cell-8	7821	100.82	99.10	0.974	337.11
Cell-9	7763	121.88	81.37	0.973	343.41
Cell-10	5377	86.82	79.40	0.959	312.80

Table 5.2: Geometrical features extracted from ALL infected segmented WBCs.

No	Area	Major axis length	Minor axis length	Solidity	Perimeter
Cell-1	15174	154.93	126.35	0.958	522.15
Cell-2	15334	150.74	140.32	0.835	674.14
Cell-3	12966	147.29	116.09	0.936	564.59
Cell-4	10075	118.12	108.75	0.974	386.07
Cell-5	10949	160.72	88.99	0.935	469.06
Cell-6	11876	147.09	107.03	0.923	569.57
Cell-7	14223	178.25	107.17	0.920	573.96
Cell-8	12678	144.62	119.35	0.880	695.47
Cell-9	14753	158.21	119.92	0.965	474.47
Cell-10	12527	144.38	113.60	0.941	494.87

5.5 Classification using ANN

The ANN classifier implementation was executed in a Matlab environment. The output for each input vector is calculated and matched with the target and error is calculated and error is propagated back by which changes are made in the weights. The default performance for feed-

forwards is mean square error (MSE). The mean MSE between the network's output vector ***a*** and the target output vector ***t*** is given by:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \quad (5.1)$$

The back propagation algorithm is used for the classification in two phases. In the first phase, the BPNN is provided with training data. Each layer of network receives the training data and propagates it to the next layers by applying the transfer function to it except the input layer. The output layer produces the vector which is compared with the training data and error is calculated. The weights of the network are updated according to the error which is propagated back to network. In the second phase, the network is tested for the classification of the problem. The test vector is fed to the network and on the basis of the output of the network, the system classifies the WBCs as lymphoblast (ALL) or normal. As explained in the previous chapter, it was the final morphological/geometric features acquired from the segmented WBCs that are used for training and testing the performance of the classifier.

5.5.1 Morphological Level Classification

Geometric features were extracted from a total of 85 segmented WBCs of which 40 were WBCs infected with ALL while the rest 45 were normal cases. The extracted geometric features were area, perimeter, Centroid, major axis length, minor axis length and solidity. These features can be used as inputs to the network for classification. The number of neurons in the input layer depends on the number of parameters/feature types extracted from the images, which is in our case 4 that corresponds to the four geometric features: *Area, Perimeter, Major Axis Length* and *Minor Axis Length*. The number of neurons in the output layer are two as there are two classes namely infected and normal (ALL and non ALL) and number of hidden layer(s) neurons are chosen manually. The default performance for feed-forwards is based on the MSE which is computed as the average squared error between the networks output and the target outputs.

There are certain parameters to vary when we check the performance of the neural network algorithm. One of these important parameters is epoch number. Figure 5.5 presents the performance of the classifier for different number of epochs. The horizontal axis shows the epoch size while the vertical axis depicts the error which is presented here in terms of cross entropy for a total of 23 epochs. It can be seen that at epoch number 17, the classifier showed its best performance which corresponds to the minimum MSE value of 0.00044549.

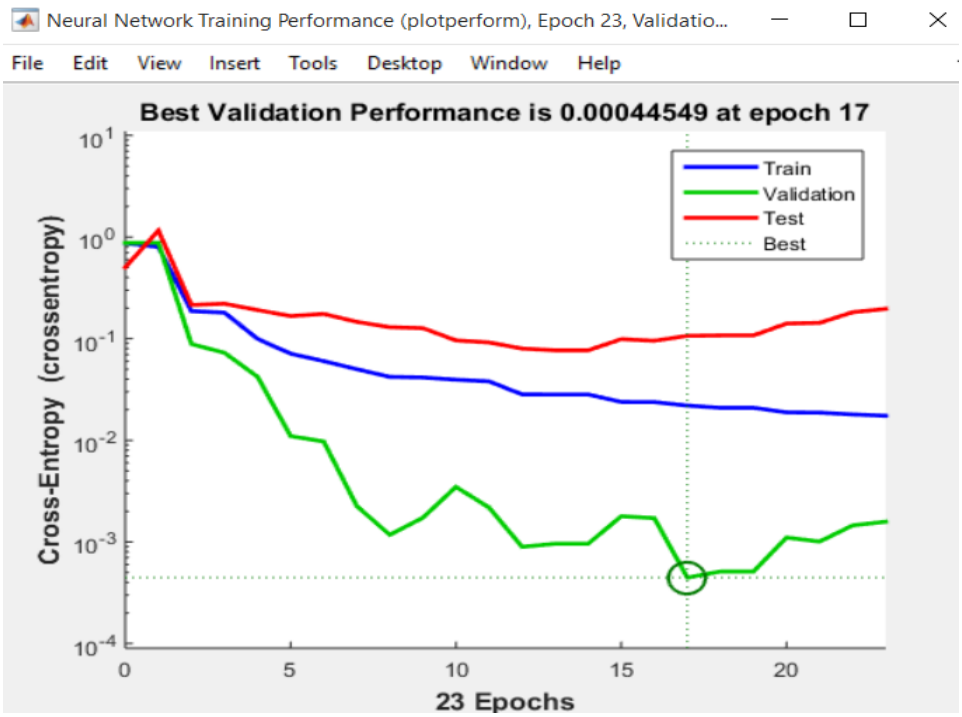


Figure 5.5: Classifier performance graph.

Figure 5.6 represents the error histogram of the trained neural network for training, validation and testing steps. The error histogram calculates the difference between the targets of our neural network and the actual outputs. The figure indicates that the data fitting errors are distributed within a reasonably good range around zero.

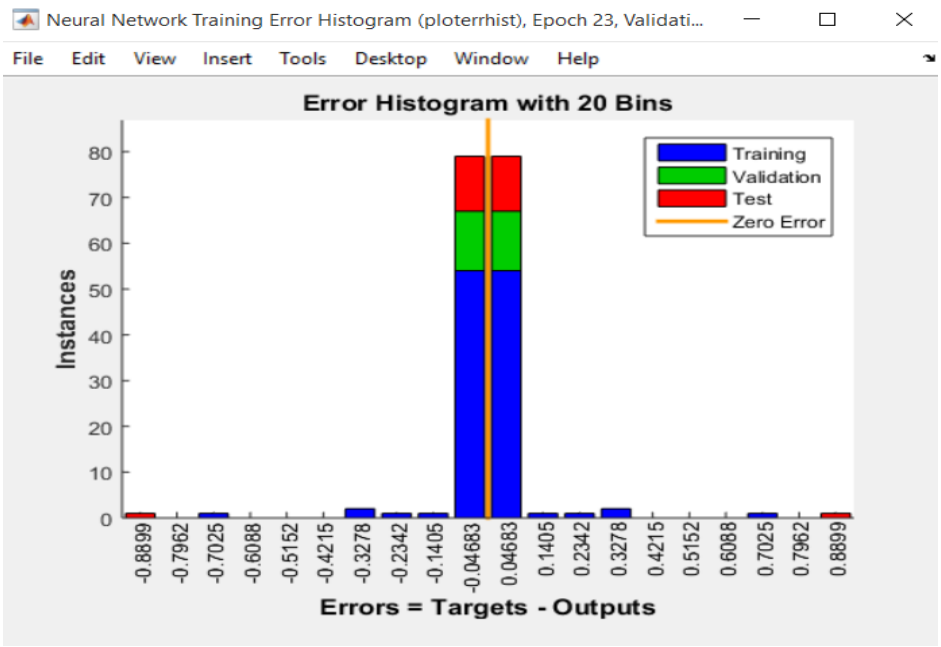


Figure 5.6: Neural network training error histogram plot.

The generated confusion matrix is shown in Fig 5.7 for training, testing and validation steps. Based on the confusion matrix, we can easily compute the specificity, sensitivity and accuracy of our algorithm which is presented in Table 5.3. Within the confusion matrix, green indicates the number of inputs assigned correctly to their classes while red indicates misclassified cases. In our case, there are two classes: the first class contains 40 ALL cases and the second class contains 45 non ALL cases. The proposed ANN classifier correctly classified 38 from the total of 40 ALL cases while all non ALL cases were correctly classified as non ALL. The black and blue cells in the confusion matrix indicate the overall results. Accordingly, the proposed scheme offered a classification sensitivity of 95%, specificity of 100% and overall accuracy of 97.6%.

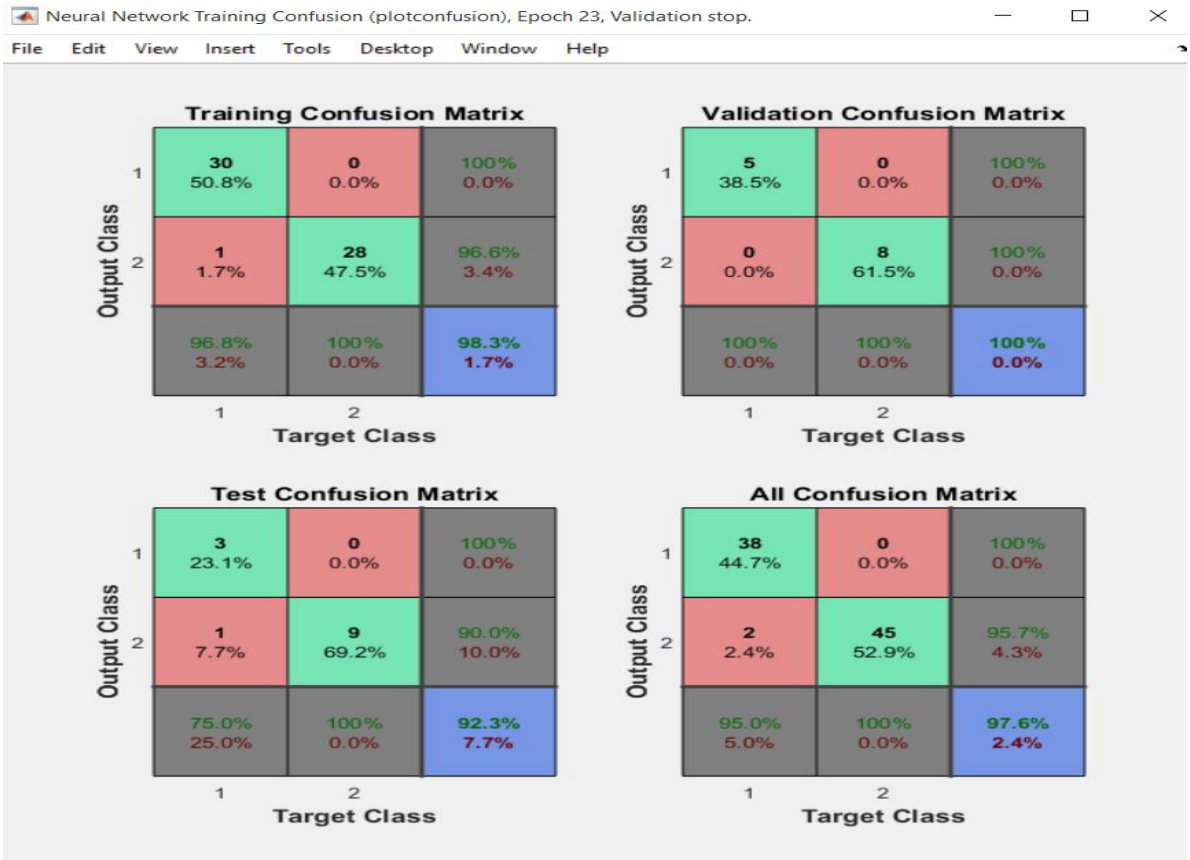


Figure 5.7: Neural network training, testing and validation performance.

Interpretation of confusion matrix of training, validation and testing data sample

The generated confusion matrix is shown in Fig 5.7 for training, testing and validation steps, the total sample are 85. The network randomly divided up 85 samples:

- 70% which is 59 samples for training, these are presented to the network during training and the network is adjusted according to its error.

- 15% which is 13 samples for validation, these are used to measure network generalization, and to halt training when generalization stops improving.
- 15% which is 13 samples for tasting, these have no effect on training and so provide on independent measures of network performance during and after training.

So those, from the fig 5.7 the training confusion matrix have contain the following parameters

TP = 30	FN=0
FP=1	TN=28

The percentage of TN $(\%\,TN) = \frac{28}{59} \times 100\% = 47.457\%$, this means 28 or 47.457% samples have ALL cases from the total 59 training samples and same as for TP, FP, and FN.

So those, from the fig 5.7 the validation confusion matrix have contain the following parameters

TP = 5	FN=0
FP=0	TN=8

The percentage of TN $(\%\,TN) = \frac{8}{13} \times 100\% = 61.538\%$, this means 8 or 61.538% samples have ALL cases from the total 13 validation samples and same as for TP, FP, and FN.

So those, from the fig 5.7 the testing confusion matrix have contain the following parameters

TP = 3	FN=0
FP=1	TN=9

The percentage of TN $(\%\,TN) = \frac{9}{13} \times 100\% = 69.23\%$, this means 9 or 69.23% samples have ALL cases from the total 13 testing samples and same as for TP, FP, and FN.

Generally the result of the classification has no any date fitting error or no data over lapping

Figure 5.8 presents the performance of the proposed classified in terms of the ROC curve. It is clear that the performance of the proposed classifier is quite promising based on all the different tests that were performed. The corresponding AUC value was 0.9787 for the final classification result.

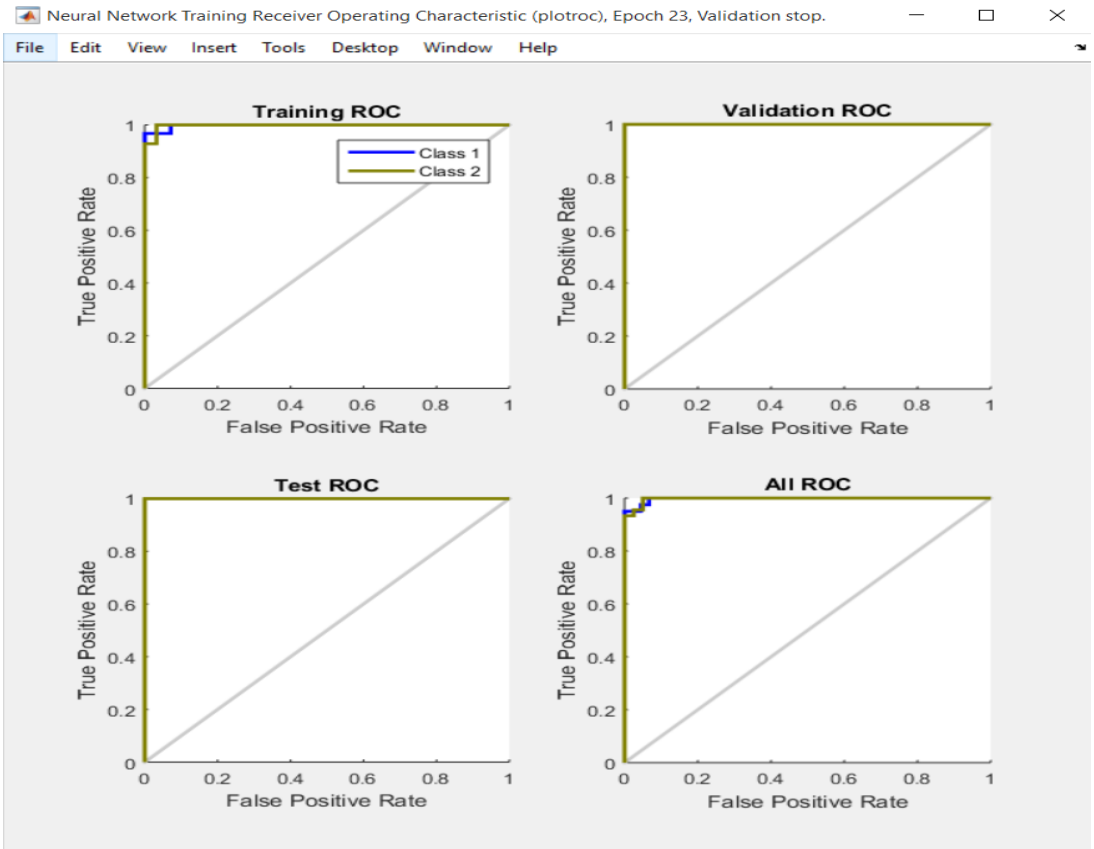


Figure 5.8: ROC curves for the neural network training, testing and validation states.

Table 5.3: Computed sensitivity, specificity and overall accuracy matrices for the proposed scheme.

No	Performance Parameter	Performance
1	Sensitivity	95.0%
2	Specificity	100 %
3	Accuracy	97.6 %

Chapter Six

5. Conclusion and Recommendations

6.1 Conclusion

Regardless of progressive techniques like patient clinical history, physical examination, and laboratory examination, microscopic examination of peripheral blood samples still remains an important screening procedure for ALL. Again it is not sufficient enough to merely make a diagnosis of ALL. Hematopathologist across the globe have been dependent on visual assessment of blood samples for diagnosis and classification of leukemia. Such human visual evaluation is time consuming, subjective and inconsistent in comparison to computerized analysis of microscopic blood images which is more accurate, rapid and quantitative. Such automation requires suitable use of image processing for improving the screening accuracy. In this thesis, attempts have been made for detecting and classifying ALL from microscopic blood images using holistic representation of the color microscopic blood image pixels in the three (trinion) space and applies trinion based integral transform (TFT) to extract useful higher order features of the RGB microscopic images after transforming them into L*a*b color space.

The major part of this work segmentation of WBCs from the complete microscopic blood image for leukemia detection. The first phase of the proposed system deals with the intensity normalization and noise removal by using image normalization and color space selection and transformation. In our case, the L*a*b color space was assumed for making the images ready for further processing. The second and major phase is leucocytes (ALL) segmentation from the L*a*b transformed microscopic images using higher order features extracted in the trinion space. Ten different types of features were extracted and among them cluster prominence feature was able to offer the best result in accurately detecting WBCs from the complete microscopic blood images. As opposed to other serial approaches that only do monochromatic processing following color component separation, holistic analysis keeps the inter-correlation information among the color bands intact. The third phase deals with image cleaning and geometric features extraction. Image cleaning was needed to neglect everything other than WBCs and partial WBCs from the image. Partial WBCs, which are often found around the edges of the blood images, must be removed in order to extract geometric features and morphological features are extracted from the cleaned signature maps. The last phase involves final classification of the blood samples into ALL infected and normal. About four geometric features were used to train and test the ANN based classification scheme: *Area, Perimeter, Major Axis Length* and *Minor Axis Length*. The model has been tasted on a total of 85 images taken under different lighting conditions and achieved an overall classification accuracy of 97.6%. It is hoped that the proposed scheme could contribute a lot in early detection of leukemia without the need for costly tests and with a better accuracy.

6.2 Recommendations

The final result of this study indicated the great promises of the proposed scheme in Leukemia studies. However, there are still rooms for improvement with the method. Firstly, the algorithm needs to be tested and validated on a larger data set and checked for its robustness. A more rigorous comparison between the proposed method and other approached suggested in the literature could then be made. Secondly the proposed algorithm has been trained and tested for only one type of Leukemia, ALL. Inclusion of the other three types of Leukemias could make the algorithm more complete. That may, however, require identification and training of more useful features that could be extracted from the microscopic sample blood images. It is believed that the idea of analyzing the color microscopic images by treating each individual pixel as a vector should set an excellent foundation for the development of such robust features that could uniquely identify the different types of Leukemias. Another possible avenue for future work could be testing efficacy of other classification schemes other than the ANN including Support Vector Machine (SVM), Genetic Algorithm (GA) and the like for their use in accurately segmenting normal and abnormal WBCs in microscopic images. These and similar other topics require further investigations.

Reference

1. E. Barnes. Disease and Human Evolution. University of New Mexico Press, 2005.
2. S. L. Gilman. Diseases and Diagnoses: The Second Age of Biology. Transaction Publishers, 2009.
3. K. Park. Textbook of Preventive and Social Medicine. Bhanot, 18th edition, 2005
4. Britt-Marie., (2003). Chemotherapy in Childhood Acute Lymphoblastic Leukemia: In Vitro Cellular Drug Resistance and Pharmacokinetics. (Doctoral Dissertation). Retrieve from: - http://uu.diva portal.org/smash/get/diva2:161996/FULLTEXT01. (24th March 2011)
5. Brunetti, A. and Haraldseth, O., (2007). Medical Imaging for Improved Patient Care. Retrieve from:http://www.esf.org/fileadmin/links/EMRC/ESF_POLICY28_V09_HD.pdf. (12th December 2010)
6. M.D., Chiu S.J., Lo J.Y., Toth C.A., Izatt J.A., Forsiu S., (2010). Novel Applications of Super-Resolution in Medical Imaging Book Chapter in Super-Resolution Imaging, Peyman Milan far (Editor). CRC Press. Page 383-412
7. Krop E.J, Acute Myelogenous Leukemia, Human Press. ISBN 15-882936-210, 2007
8. Robison. M.D, Chiu S.J., Lo J.Y., Toth C.A., Izatt J.A., Forsiu S. *Novel Applications of Super-Resolution in Medical Imaging* Book Chapter in Super-Resolution Imaging, Peyman Milanfar (Editor), CRC Press, Page 383-412, 2010).
9. Biondi, A., Cimino, G., Pieters,R., Pui, C. H., Biological and Therapeutic Aspects of Infant Leukemia, 2000
10. Marirb Katja, and Elaine Nicpon. Human anatomy and Physiology personal Education Publishing, 9th edition, 2013.
11. J.Bain. a Beginners Guide to Blood cells. Blackwell Publishing, 2nd edition,2004
12. A.Scott and E.Fong. Body Structure and Functions. Thomson, 10th edition, 2004
13. S.Beeker. A.Handbook of Chinese Hematology. Blue Poppy Press, 2000
14. G.E.Xueling and X.Wang. Role of canonical pathway in hematological malignancies. Journal of Hematology and Oncology,3(33):1756-8722,2010
15. Sisay. Y and Dagnachew.M. Childhood cancer in Gonder University Hospital, Northwest Ethiopia, Published online 2015 sept 24.
16. B .Leonard, editor. Leukemia: A Research Report. Diane, 1993.
17. A. Bordoni D.R. Abreul and E. Zucca. Epidemiology of hematological malignancies. Annals of Oncology, 18(1):3-8, 2007.
18. M.A. Lichtman, E. Beutler, T.J. Kipps, U. Seligsohn, K. Kaushansky, and J.T. Prchal, editors. Williams Hematology. McGraw Hill, 2007.
19. D. Wartenberg, F.D. Groves, and A.S. Adelman. Acute lymphoblastic leukemia: Epidemiology and Etiology. In Acute Leukemias, Hematologic Malignancies, pages77-93. Springer Berlin Heidelberg, 2008.

20. D.A. Casciato and M.C. Territo, editors, Manual of Clinical Oncology. Lippincott Williams and Wilkins, 6th editions, 2004.
21. L.J. Kinlen. Epideminological evidence for an infective basis in childhood leukemia. British Journal of cancer, 71(1):1-5, 1995.
22. D.L. Preston, S. Kusumi, M. Tomonag, S. Izumi, E. Ron, A. kuramoto, N. Kamada, H. Dohy, and T. Matsuo. Cancer incidence in atomic bomb survivors. Part III: Leukemia, Lymphoma and Multiple myeloma, 1950-1987. Radiation Research, 137(2): 68-97, 1994.
23. A. Stewart and R. Barber. A survey of childhood malignancies. British Medical Journal, 28: 1497-1958.
24. D.M. Pelissari, F.E. Barbieri, and F. V. Wnsch. Magnetic fields and acute lymphoblastic leukemia in children: a systematic review of case-control studies. Cadernos de saude publica 25: S441-S452, 01 2009.
25. M.A.C. Jimnez and L.O. Vargas. Parental exposure to carcinogents and risk for Childhood acute lymphoblastic leukemia, Colombia. Preventive Chronic Disease, 8(5), 2011.
26. O.P. Soldin, H.N. Maktabi, J.M. Genkinger, C.A. Loffredo, J.A.O. Garcia, D. Colantino, D.B. Barr, N.L. Luban, A.T. Shad, and D.Nelson. Pediatric acute lymphoblastic leukemia and exposure to pesticides. Therapeutic Drug Monitoring, 31(4), 2009.
27. T. Singh. Textbook of Hematology. Arya Publication. 2004.
28. W.Wen, X.O. Shu, J.D. Potter, R.K. Severson, J.D. Buckley, G.H. Reaman, and L.L. Rebison. Parental medication use and risk of childhood acute lymphoblastic leukemia. Cancer, 95(8):1786-1794, 2002.
29. Howell DA, Smith AG, Jack A, et al. Time-to-diagnosis and symptoms of myeloma, lymphomas and leukaemias: a report from the Haematological Malignancy Research Network. BMC Hematol. 2013; 13(1):9. [PMC free article] [PubMed
30. A.K. Agarwal, editer. Clinical Medicine: A practical manual for students and practitionars. Jaypee, 2007.
31. T. Singh. Atlas and Text of Hematology. Avichal, 2010.
32. T. M. Elsheikh, S. L. Asa, J. K. Chan, R. A. DeLellis, C. S. Heffess, V. A. LiVolsi, and B. M. Wenig. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. American Journal of Clinical Pathology, 130(5):736–744, 2008.
33. L.A.D. Cooper, A.B. Carter, A.B. Farris, W. Fusheng, K.Jun, D.A. Gutman, P. Widener, T.C. Pan, S.R. Cholleti, A. Sharma, T.M. Kurc, D.J Bart, and J.H. Saltz. Digital Pathology: Data-intensive frontier in medical imaging. Proceedings of the IEEE, 100(4):991-1003,2012
34. M.E Tathagata Rav, D.S Reddy, A. Mukherjee, J. Chatterjee, R.R. Paul, and P.K. Dutta. Detection of constituent layers of histological oral sub-mucous fibrosis: Images using the hybrid segmentation algorithm. Oral Oncology,44(12):1167-1171,2008

35. M. Dong, M. Eramian, S.A.Ludwing, and Roger A. Pierson. Automatic detection and segmentation of bovine corpora lutea in ultrasonographic ovarian images using genetic programming and rotation invariant local binary patterns. Medical and Biological Engineering and Computing, 51(40:505-416,2013
36. S.J. Keenan, J. Diamond, W.G.McCluggage, H.Bharucha, D.Thompson, P.H. Bartels, and P.W. Hamilton. An automated machine vision system system for the histological grading of cervical interaepithelial neoplasia. The Journal of Pathology, 192(3):351-362, 2000
37. P. Khurd, C. Bahlmann, P.Maday, A.kamen, S.Gibbs-Strauss, E.M. Genega, and J.V. Frangioni. Computer-aided gleason grading of prostate cancer histopathological imaging using texton forests. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 636-639, 2010.
38. A.N. Basavanhally, S. Ganesan, S. Agner, J.P. Monaco, M.D. Feldmanad, J.E. Tomaszewski, G.Bhanot, and A. Madabhushi. Computerizes image-based detection and grading of lymphocytic infiltration in her2+breast cancer histopathology IEEE Transactions on Biomedical Engineering, 57(3):642-653, 2010.
39. E. Ozdemie, C. Sokmensure, and C. Gunduz Demire. A resampling- based markovian model for automated colon cancer diagnosis, IEEE Transactions on Biomedical Engineering, 50(1):281-289, 2012.
40. K.Belkacem-Boussain, M.Pennel, G.Lozanski, A.Shanaah, and M.Gurcan. Computer-aided classification of centroblast cells in follicular lymphoma. Analytical and Quantitative Cytology and Histilogy, 32(5):254-260, 2010.
41. Gonzalez, R.C, and Woods, R. image processing. 2nd Edition Prentic hall. ISBN-0-130-94650-8, 2002.
42. M. Rangayyan, Begona Acha, Carmen Serrano. Color image processing with biomedical application. Society of Photo-Optical Instrumentation Engineering (SPIE), USA, 2011.
43. "RGB color space," [Online].Available: https://en.wikipedia.org/wiki/RGB_color_model [October 18, 2016]
44. K.A. Thomas. Image processing as applied to medical diagnostic. MSC dissertation, university of Oregon, 2010.
45. H. D. Cheng, X. H. Jiang, Y. Sun, and J. L. Wang, Color image segmentation: advances and prospects, Pattern Recognition, 34(12), pp. 2259-2281, 2001.
46. C. H. Li and T. Yuen, Regularized color clustering in medical image database, IEEE Transaction on Medical Imaging, 19(11), pp. 1150-1155, 2000.
47. T. Acharya and A.K. Ray. Image Processsing Principles and Application. Wiley-Interscience, 2005.
48. Basavaprased B, Ravi M. A comparative study on classification of image segmentation methods with a focus on graph based techniques. IJRET, vol.03, 2014.
49. C.Cliu and P.Chung. Object extraction algorithm of color image. ICIC, vol.7, no 10, pp.5771-5787, 2011.

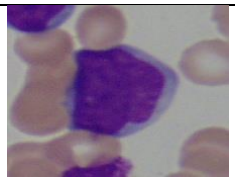
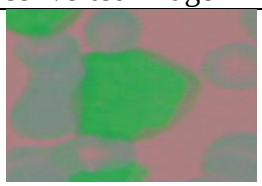
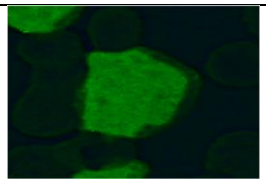
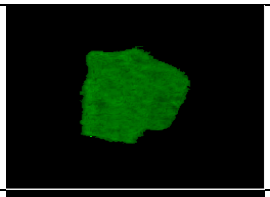
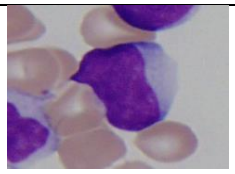
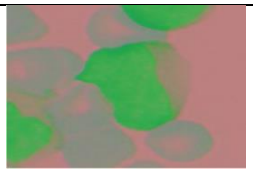
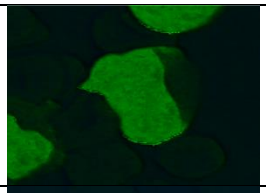
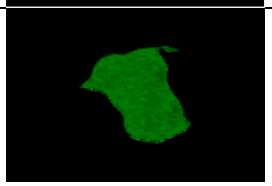
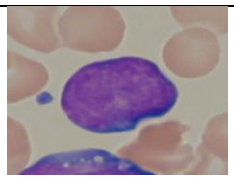
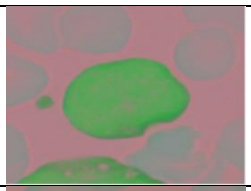
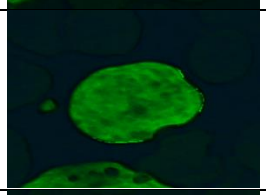
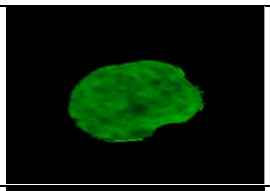
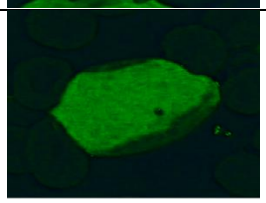
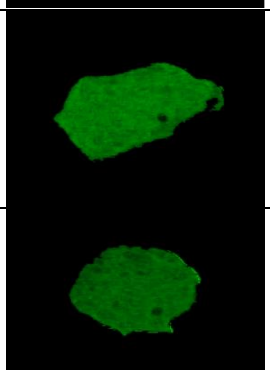
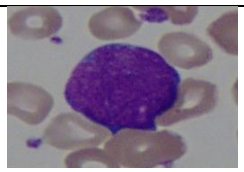
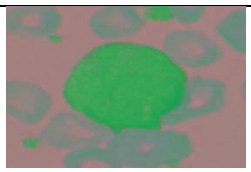
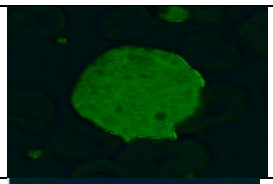
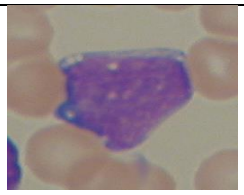
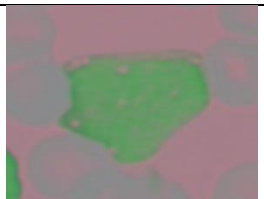
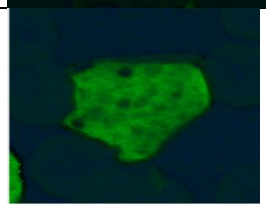
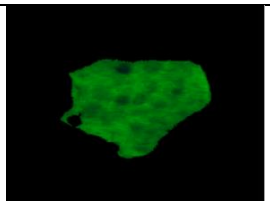
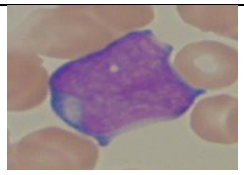
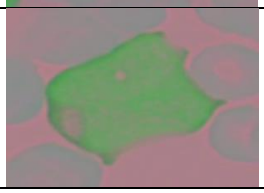
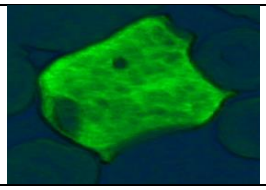
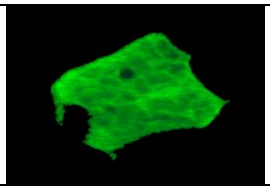
50. Huang, Zhi-Kai, and Kwok-Wing Chau. A New Image Thresholding Method Based on Gaussian Mixture. Applied Mathematics and Computation. 899-907, 2008.
51. Morse, B.S. lecture 4: Thresholding. Retrieved from: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL COPIES/MORSE/th reshold.pdf, 1st April 2010.
52. Madhloom HT, Kareem SA, Ariffin.H, Zaidan. AA, Alanazi.HO, Zaidan BB. An automated white blood cell nucleus localization and segmentation using image arithmetic and automated threshold. J Appl sci 2010; 10(11):959-66.
53. Q. Liao and Y. Deng. An accurate segmentation method for white blood cells images. In Proceedings of the IEEE International Symposium on Biomedical Imaging, pages 245-248, 2002.
54. Halium 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))
55. Tytuiu97iujkl
56. Subrjee Mohapatra, Dipi Patra, Sanghmitra Satpathy, Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection Using Color Based Clustering. International Journal of Computer Information System and Industrial Management Application, vol4, pp.477-485, 2012.
57. S. Chinwaraphat, A. Sanpanich, C. Pintavirooj, M. Sangworasil, and P. Tosranon. A modified fuzzy clustering for white blood cells segmentation. In Proceedings of the Third International Symposium on Biomedical Engineering, volume 6, pages 2259-2261,2008.
58. N.T. Umpon. Patch based white blood cell nucleus segmentation using fuzzy clustering. ECTI Transaction Electrical Electronics Communications, 3(1):5-10, 2005.
59. 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.
60. 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 morphometric. In Proceedings of the IEEE Students Technology Symposium, pages 59-66, April 2010.
61. 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.
62. L.Gupta, S.Jayavanth, and A. Ramaia. Identification of Different types of Lymphoblast in Acute Lymphoblastic leukemia Using Relevance Vector Machines. In Proceedings in Medicine and Biology Society, USA, 2009.
63. R. Duda, D. Hart, and P. Stork. Pattern Classification. Wiley India, 2nd edition, 2007.
64. T. Acharya and A. K. Ray. Image Processing Principles and Applications. Wiley-Interscience, 2005.
65. G.D. Tourassi, Journey toward computer-aided diagnosis: Role of image texture analysis, Radiology, pp317-320, 1999.

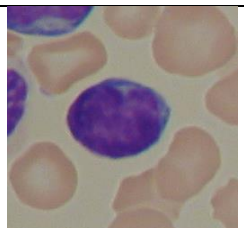
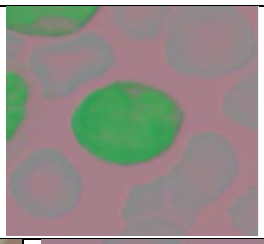
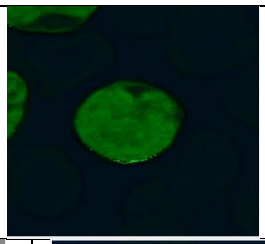
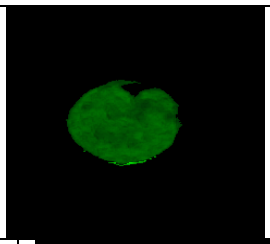
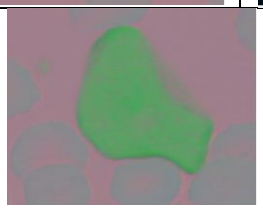
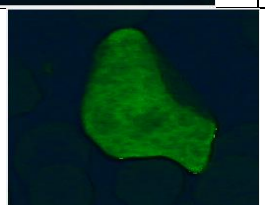
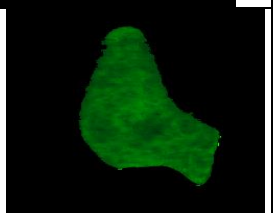
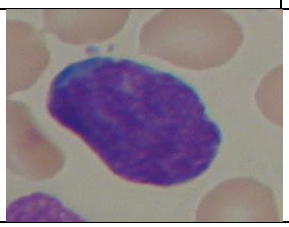
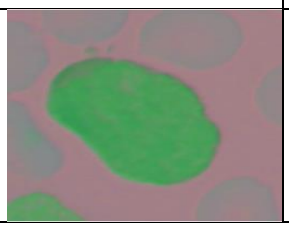
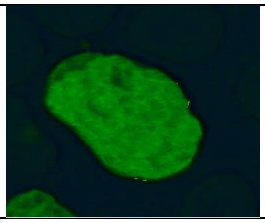
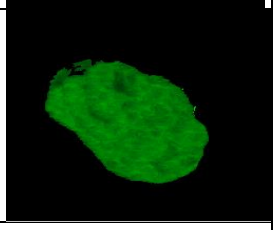
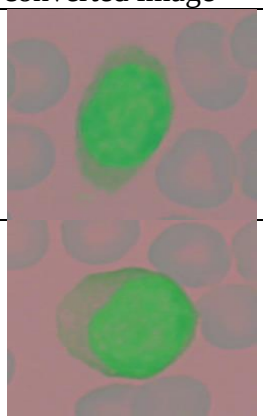
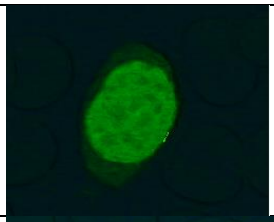
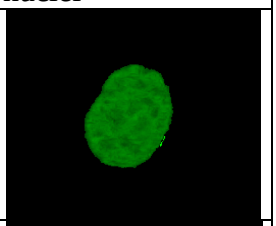
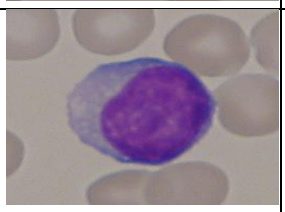
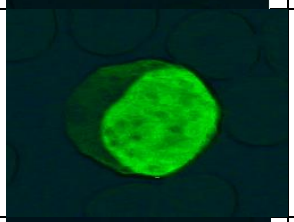
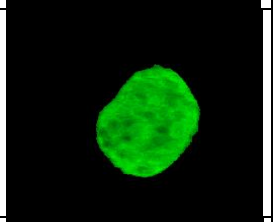
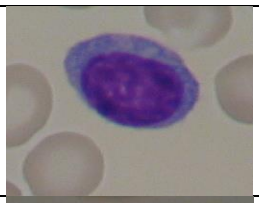
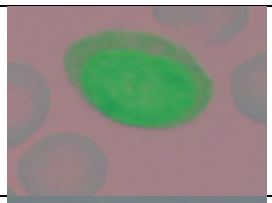
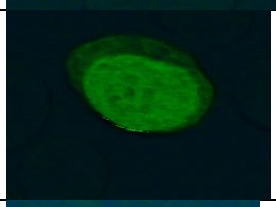
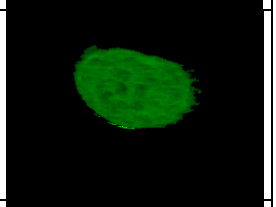
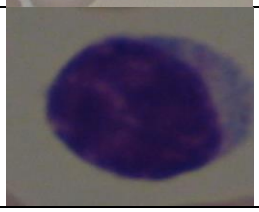
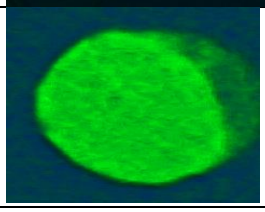
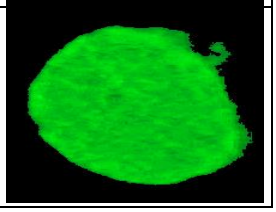
66. Bo Li, Wei Li, and Dazhe Zhao. Global and Local Features Based Medical Image Classification. J.Med.Imaging Inf.5, pp.748-754, 2015.
67. Nidhal K. Al Abbadi et al. Psoriasis Detection Using Skin Color and texture features. J. computer Sci, vol.6 no.6.pp.648-652, 2010.
68. D. Ubeyli and I. Guler. Multilayer perceptron neural networks to compute quasistatic parameters of asymmetric coplanar waveguides. Neurocomputing, 62:349 – 365, 2004
69. Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal, "A Survey on Image Classification Approaches and Techniques," IJARCCE, vol.2, pp.1005-1009, 2013.
70. Joaquim Jose Furtad, Zhihua Cai & Liu Xiaobo, "Digital image processing: supervised classification using genetic algorithms," Report and Opinion, vol.2, no.6, pp.53-61, 2010.
71. R. Duda, D.Hart, and P. stork. Pattern Classification. Wiley India, 2nd edition, 2007.
72. E. D. Ubeyli and I. Guler. Multilayer perceptron neural networks to compute quasistatic parameters of asymmetric coplanar waveguides. *Neurocomputing*, 62:349 – 365, 2004.
73. Simon Haykin. Neural Networks. Prentice Hall, 2nd edition, 1999.
74. V. N. Vapnik. Statistical Learning Theory. John Wiley and Sons, 1998.
75. R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in 2011 18th IEEE International Conference on Image Processing, 2011, pp. 2045–2048.I
76. D. Assefa, L. Mansinha, K. F. Tiampo, H. Rasmussen, and K. Abdella, Local quaternion Fourier transform and color image texture analysis, Sig Proc., 90(6), pp. 1825-1835, 2010.
77. D. Assefa, L. Mansinha, K. F. Tiampo, H. Rasmussen, and K. Abdella, The trinion Fourier transform of color images, Sig Proc., 91(8), pp. 1887-1900, 2011.
78. Dawit Assefa, H. Keller, D. A. Jaffray, "Signal analysis of multi-parametric MR images in higher order Fourier spaces," International Journal of Computational Bioscience, vol.4, no.1, 2013.
79. D. Moges, B. Assefa, Dawit Assefa, "A mathematical algorithm for robust color retinal image analysis - a preliminary study," 7th Ethiopian ICT Annual Conference, EICTAC2014, Addis Ababa, Ethiopia, June 6 2014.
80. X. Gou, Z. Liu, W. Liu, Y. Xu, "Three-dimensional wind profile prediction with trinionvalued adaptive algorithms," *IEEE Int Conf on Digital Signal Processing, Singapore*, July 2015.
81. Dawit Assefa, Ondrej Krejcar, "Novel Edge Detection Scheme in the Trinion Space for Use in Medical Images with Multiple Components," 8th International Conference on Computational Collective Intelligence (ICCCI 2016), Halkidiki, Greece, 28th-30th September 2016, LNCS.
82. R. M. Haralick, "Texture Features for image classification," IEEE Transactions on systems, Man and Cybernetics, vol. SMC-3, no. 6, 1973.
83. Russel, S. and Norvig.P. Artificial Intelligence: A Modern Approach. ISBN 0-13-080302, 2003.

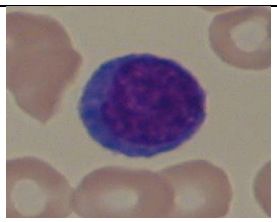
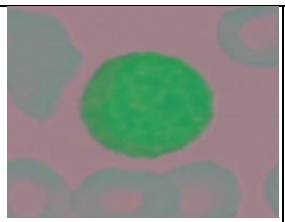
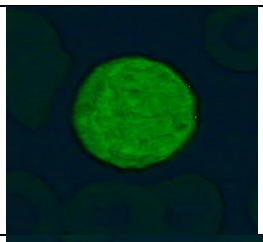
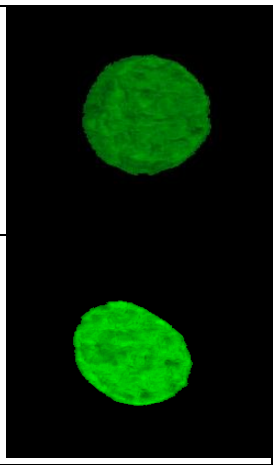
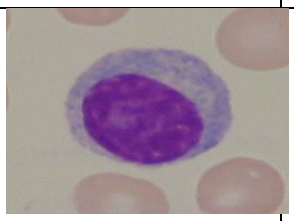
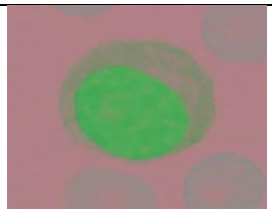
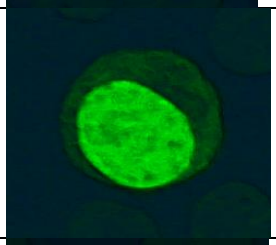
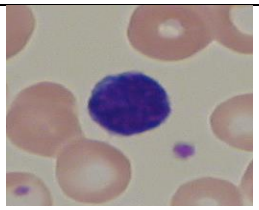
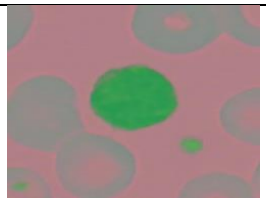
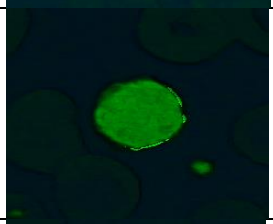
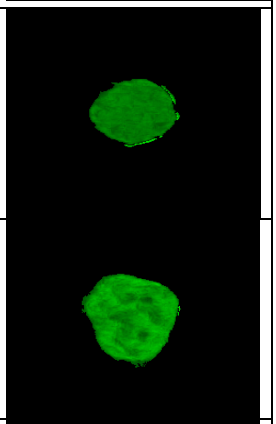
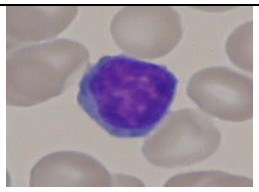
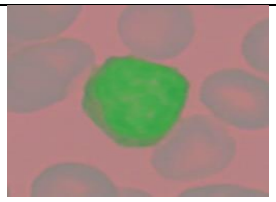
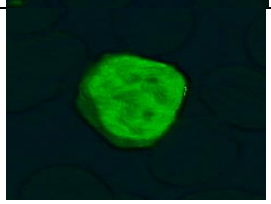
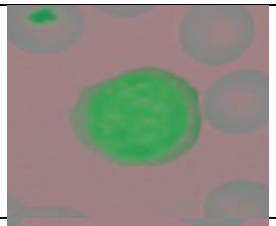
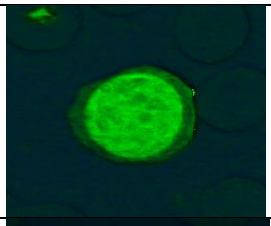
84. The Nervous system. Retrieved from: http://WWW.naturalhealthschool.com/9-2.htm. 14th march 2011.
85. Gurney, K. An Introduction to Neural Networks. UCL Press Limited. ISBN 1-85728-673-1, 1997.
86. Freeman.J.A and Skapura D.M. Neural Networks: Algorithms, Applications, and programming Techniques. Addison-Wesley Publishing Company. ISBN 0-201-51376-5, 1992.
87. Picton, P. Neural Network. Palgrave. ISBN 0-333-80287-X, 1994.

Appendix

Results of the TFT based Segmentation

Leukemic WBC segmentation				
No	Original image	Lab colorspace converted image	Signature map	Final segmented nuclei
Cell -1				
Cell -2				
Cell -3				
Cell -4				
Cell -5				
Cell -6				
Cell -7				

Cell -8				
Cell-9				
Cell-10				
Normal WBC segmentation				
No	Original image	Lab colorspace converted image	Signature map	Final segmented nuclei
Cell-1				
Cell-2				
Cell-3				
Cell-4				

Cell-4				
Cell-5				
Cell-6				
Cell-7				
Cell-8				
Cell-9