Myeloma Cell Detection in Bone Marrow Aspiration Using

Microscopic Images

by

Muhammad Sikander Zamir

01-243152-007

Supervised By

Dr. Samabia Tehsin

A thesis submitted in partial fulfillment for the degree of

Master of Science in Computer Science

in the

Department of Computer Science

Bahria University, Islamabad

20187

Abstract

Multiple myeloma is a haematological cancer that occurs in plasma cells. The increase in the myeloma cells induces the reduction in all the blood cells. Production of abnormal antibodies or multiple myeloma proteins (for example, IgG or IgA) is the consequence of multiple myeloma cells.

Myeloma cells in bone marrow images are detected automatically in the current study. Various steps including preprocessing, feature extraction and segmentation are used to detect myeloma cells in bone marrow because microscopic examination is not sufficient enough. This inadequacy is because of involvement of human factors i.e., fatigue, stress, proficiency, experience etc. Therefore, there is utmost need of some automated method to detect myeloma cells in bone marrow aspirations to achieve accuracy in performance, and making the entire process time-effective.

Convolutional neural network is employed in order to detect multiple myeloma while using microscopic images of blood. Pre- trained and fine- tuned AlexNet is taken into consideration instead of developing an architecture from scratch. Input images are classified into normal and blast by substituting last layer of pre- trained AlexNet by two new layers. Besides, there is the need of classifier in order to classify myeloma cells. Hence, SVM is used for this purpose.

Purposed methodology was compared with the previous literature in terms of total number of images in the data set and their performance as well. Our methodology unlike previous ones performed well in terms of carrying out 100 % accuracy without employing any microscopic segmentation.

Contents

List of Figures

List of Tables

Abbreviations

ACC Accuracy

ACIS Automated Cellular Imaging System

ALL Acute Lymphoblastic Leukemia

ANN Artificial Neural Networks

BMC Bone Marrow Cells

BMPC Bone Marrow Plasma Cell

CAD Computer Aided Diagnose

CNN Convolutional Neural Network

CT Computed Tomography

FCM Fuzzy C-means

FITC Fluorescein Isothiocyanate

F_N False Negative

F_P False Positive

GLCM Gray Level Co–occurrence Matrices

ISH In Situ Hybridization

KNN K- nearest neighbour

LVQ Learning Vector Quantization

MLP Multi-Layer Perceptron

MM Multiple Myeloma

MRI Magnetic Resonance Imaging

PE Probability of Error

PET Positron Emission Tomographies

RBCs Red Blood Cells

ReLu Rectified Linear units

RF Random Forest

SAM Stepwise Averaging Method

SC Specificity

SS Sensitivity

SVM Support Vector Machine

T$_N$ True Negative

T$_P$ True Positive

WBCs White Blood Cells

CHAPTER # 1

Introduction

1.1   Blood

Blood is composed of two components, plasma and cells namely erythrocytes, leucocytes, and thrombocytes. Erythrocytes, leucocytes, and thrombocytes are also termed as RBCs, WBCs, and platelets respectively. RBCs contain an iron containing pigment haemoglobin that carries oxygen from lungs. The main role of WBCs is to fight against infections and henceforth play a major role in the formation of antibodies (immunoglobulins) in response to foreign particles, which are crucial for body's defence system. In addition, leucocytes also include lymphocytes that play major roles in the defence system. Basically, leucocytes are assembled in to myeloid and lymphoid stem cells [1]. Moreover, thrombocytes play an important function in blood clotting.

1.2   Formation of Blood

Furthermore, the formation of all blood cells takes place in bone marrow [2] which is the centre of production of majority of the cells circulating in the blood. It, therefore, comprises of these cells as well as their precursors. The white blood cell lineage consists of granulocytes, lymphocytes and monocytes. Lymphocytes include B-lymphocytes which give rise to plasma cells. The entrance of bacteria or viruses in the body accounts changes in the B cells into plasma cells. Plasma cells are responsible for producing antibodies that fight against the bacteria and viruses and hence prevent any disease and infection [3].

1.3   Multiple Myeloma

Furthermore, uncontrolled production of plasma cells is classified as neoplasia that can be either benign or malignant. Characteristic of multiple myeloma is the accumulation of malignant plasma cells in the bone marrow [5]. Multiple myeloma is a haematological cancer that occurs in plasma cells. The increase in the myeloma cells induces the reduction in all the blood cells.

Production of abnormal antibodies or Multiple myeloma proteins (for example, IgG or IgA) is the consequence of multiple myeloma cells. One of the other features of multiple myeloma is increase in the level of M protein in the blood.

Moreover, each myeloma cell is similar to one another. However, M proteins provide no advantage to an individual. The increase in level of M proteins also disturbs normal functioning immunoglobulins which ultimately causes damage to bones or sometimes renal complications [5]. Multiple myeloma is commonly accompanied with bone diseases comprising osteolytic bone lesions, osteoporosis, pathological fractures, and hypercalcaemia. Besides, bone marrow cells damage the hard layer of the bone because of myeloma cells. This results in softening of bones or osteoporosis which eventually increases the danger of bone injuries as multiple myeloma is characterized by the high capacity for bone destruction [6]. Almost all patients with MM have early osteolytic lesions due to increased bone contraction related to stimulation of osteoclastic recruitment and activity in the direct locality of myeloma cells. On the contrary, symptoms of osteolytic lesions are not present in every patient suffering from multiple myeloma. [5]

1.4   Symptoms of Multiple Myeloma

The studies of Lahtinen *et al* (1992) [7] shows that 50% of patients have spinal breakage, while 30% have non-spinal breakage. The reasons behind the production of osteolytic lesions in multiple myeloma are unclear. Whereas the studies of Fleisch (1991) indicate that medicines like bisphonsphonates are proven to be beneficial in the treatment of bone resorption [8]. Current researches have stated that bisphonsphonates have been taken in the treatment of multiple myeloma on large scale [7, 9]. Another drug clodronate serves to lessen bone fractures [7] and therefore minify spinal and non- spinal breakages [9]. Further, McCloskey *et al* (1996) studies proved that clodronate could raise the mortality rate in patients having no spinal breakages [9].

Multiple Myeloma results in nearly 1% of all cancers and accounts 2% of death-rate. Multiple myeloma commonly occurs at the age of 65 to 70 years, but current figures designate that it can also occur at early age of life. Multiple Myeloma is usually considered fatal but highly treatable. In order to differentiate Multiple Myeloma from other plasma cell dyscrasias, the diagnosis is

done on the basis of cytologic and histologic features. "A bone marrow aspiration is usually performed to estimate the percentage of bone marrow occupied by plasma cells which is used in the diagnostic criteria for myeloma" [10].

1.5   Identification of Multiple Myeloma

The identification of MM and dyscrasias of plasma cell is done through the microscopic investigation of the bone marrow and monitoring theory which plays a fundamental role in the whole procedure. Most importantly, the evidence about the level of bone marrow association with plasma cells disorders is provided by this. The identification of multiple myeloma is done by determining the plasma cell percentage in the bone narrow.

It is hard to identify myeloma by merely doing a single laboratory test so it is identified by doing many other tests. Some of the factors that are involved in the perfect identification of myeloma are the history of patient, body check-up, signs, and results of laboratory tests. The fundamental tests consist of blood and urine tests along with the biopsy of bone marrow. X-rays, magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomographies (PET) are some other tests. Change in the structure of bone along with mass and content of MM tumours in the bone is acknowledged by these tests.

In a biopsy of bone marrow, a small piece of bone that contains marrow is removed. For a bone marrow aspiration, a slight quantity of liquid bone marrow having bone marrow cells is removed. A large needle is used to take sample from both pelvic and hip bone. The amount of normal plasma cells and multiple myeloma cells in the bone marrow is determined through this test.

The bone marrow biopsy is done to diagnose the problems associated with erythrocytes, leucocytes, and thrombocytes such as anemia or high or low count of white blood cells in body, leukemia or multiple myeloma. Multiple myeloma is confirmed through the existence of myeloma cells in the bone marrow. However, severity is determined on the basis of large proportion of myeloma cells [11].

1.6   Diagnosis of Multiple Myeloma

According to Smith, A, *et al.* (2005)*,* initial investigation of myeloma patients should include the screening tests including FBC, ESR or plasma viscosity, serum or plasma electrolytes, urea, creatinine, calcium, albumin and uric acid, electrophoresis of serum and concentrated urine, quantification of non-isotopic immunoglobulins, and X-ray of symptomatic areas, should be implemented. In addition, immunofixation confirms the presence of M-protein, as immunofixation strongly indicates multiple myeloma and it is the demonstration of M-protein that confirms multiple myeloma. [12].

Pathologists generally detect myeloma cells by implementing microscopic techniques. Staining is used to make bone marrow smears for microscopic identification. All blood cells along with many artefacts are stained by using Giemsa. Giemsa changes the colour of cytoplasm from pink to blue and the nuclear material from red to purple. Death rates are reduced through early diagnosis and treatment of multiple myeloma. The identification of multiple myeloma is done by using microscope to find myeloma cells in the bone marrow slides. Though, microscope observation results in delayed and flawed identification of multiple myeloma. Hence computer mediated techniques including image based identification methods can be used to overthrown the problem associated with manual microscopy. Recent advancements in image processing techniques assist the haematologists to automatically identify the diseases through microscopic images. This has proved to be constructive even for those who are not properly trained in the identification of multiple myeloma [18].

Since not much work has been done in this field our research could prove an important step in the myeloma diagnostic process.

1.7   Research Problem

"Identification of Myeloma cell in microscopic bone marrow aspiration images"

In 2015 the annual death toll of multiple myeloma was 101,100 amongst the 488,000 affected globally. According to the U.S statistics, it affects 6.5 per 100,000 annually. The occurrence is more common in men than women and usually at the age of 61. With treatment, the survival can

be up to 4-5 years, without which it is reduced to seven months on average. There are several promising treatments and therapies available all of which demand early diagnosis. Manual diagnosis comes with the disadvantage if poor accuracy, time consumption and a greater risk factor. Therefore, we have proposed an effective automated system which will process bone marrow smear images. It will separate RBCs from WBCs and help in the identification of myeloma cells. This methodology will considerably improve speed and accuracy of the diagnosis.

1.8   Objectives

Following are the research objectives for the proposed research question.

- Identification of white blood cells (WBCs)
- No segmentation as required in standard methods
- Classification of myeloma and non-myeloma cell
- Effective automated system for myeloma
- Higher accuracy as compared to all previous methods

CHAPTER # 2

Literature Review

In the previous literature there is lot of work done for the classification of white blood cells (WBCs). However myeloma cell detection is neglected.

Smith, A, *et al.* have proposed a methodology for myeloma cell detection. The method is started by some pre-processing steps and the main algorithm is based on cell detection and cell splitting carried out using bottleneck algorithm, modified watershed and SVM classifiers to detect myeloma cells [12].

Saeedizadeh, Zahra, *et al* have proposed the computer mediated diagnostic procedures to diagnose multiple myeloma. Feature extraction along with other standard rules proved to be significant in identifying plasma cells. Classifier was used to distinguish between normal plasma cells and myeloma cells. In addition, the application of computer mediated aids to determine myeloma cells showed the sensitivity of 96.52%, specificity of 93.04%, and precision of 95.28%. Before the pre-processing method, the colour and brightness of erythrocytes and cytoplasm was somehow very close and incomprehensible. So after applying algorithm enhancement, erythrocytes and cytoplasm got separated from each other. Though, this caused the reduction in the algorithm accuracy. The algorithm possesses the capacity of detecting accurate amount of myeloma cells, and it is this capacity that makes the possibility of constructive treatment of multiple myeloma. On the other hand, its identification is laborious and error may take place due to the dependency on the decision of human eye and suggestions. Henceforth computer based identifications techniques are employed to lessen the time [13].

Sarrafzadeh *et al*. compared clump splitting via bottleneck detection and shape classification to the state-of-the-art algorithms for the clump splitting problem and proved to be superior to the later [14].

Mokhtar *et al.* have assessed the accuracy of SVM classifiers based on different morphological features of white blood cells using geometrical features, statistical features and moment invariants. Image enhancement techniques using local, global, bright, dark and partial contrast

stretching for acute leukaemia images were compared in which partial contrast stretching proved to be the most effective. The classifiers proved to be 93% accurate [15].

Claire, M. has reported that bisphosphonates are used to prevent bone lesions resulting from different diseases including multiple myeloma as the treatment through it is specifically accompanied with the survival of patients of this disease. The authors have investigated whether the bisphosphonates clodronate, pamidronate, and YM175 can directly affect the human myeloma cell lines U266-B1, JJN-3 and HS-Sultan in *vitro*. The investigated data revealed that some of the bisphosphonates could have direct anti- tumour effects on multiple myeloma cells in *vivo* [17].

According to Samer, Z., [18], bone marrow assessment is an important feature in the diagnosis of patients with multiple myeloma. The calculating of bone marrow aspirates using different microscopic techniques is the most important process to assess bone marrow plasma cell percentages. However, this calculating is not properly assessed. Hence CD138 is used to evaluate plasma cells, since it allows excellent evaluation of plasma cell. The authors have estimated the plasma cell percentages in 79 bone marrows from patients having multiple myeloma. The use of CD138 to determine plasma cell percentage revealed the benefits of computer-assisted image analysis. They concluded that investigation of CD138 sections should be thought of as a common method for estimating plasma cells in the bone marrow. [18]

Rajkumar et.al [19] has discussed the methods for estimating bone marrow plasma cells that are involved in myeloma. The study was conducted to discern the exemplary method identify the plasma cells of bone marrow that are involved in myeloma. Authors have studied seventy five patients including fifty men and twenty five women for the purpose of their study. Biopsy, aspirate, and PCLI are used for determining BMPC percentage. The results have shown that highest percentage of BMPC is an ideal method of identifying the patients that have high number of bone marrow plasma cells.

Patel and Mishra [20] have studied the leukaemia identification by using the automated approach. The purpose of their study is to detect the leukaemia via automated approach since manual procedure is lengthy and time-consuming as well. Moreover, there is lack of accuracy because of the involvement of human factor. Henceforth, automated system is applied on

microscopic images to overcome the lacking of manual process. It includes the application of featuring techniques and extraction of required parts of images. For classification, SVM is used. The procedure was experienced on image dataset and executed on MATLAB. Accuracy of 93.57% was achieved via this system.

Cao, Liu, and Song [21] have proposed a novel procedure to analyse microscopic images to detect bone marrow cell. In the study, bone marrow cells were classified various techniques including SVM along with Artificial Neural Networks, Random Forests, Bayesian and Adaboost networks. The accuracy of classification via SVM is better than four other classifiers i.e., 87.49%. Thus, SVM was adopted for detecting bone marrow cells (BMC).

Brown et. al [22] have detected myeloma cells through mRNA in situ hybridization (ISH). They have conducted longitudinal research for 2- 3 years on 7 patients. Moreover, the studies were conducted on 83 samples of those patients. This specific technique is most widely used to determine and monitor myeloma cells in the blood of patients who have myeloma. The quantity of circulating myeloma ranged from 0.1 to 23% of mononuclear cells, thus endorsing clinical and laboratory state of the patients.

Rawatet.al [23] have identified leukaemia by using microscopic images via Computer Aided Diagnostic system (CAD). The study was carried out to distinguish acute lymphoblastic leukemia (ALL) from lymphocytes of healthy persons. In Computer Aided Diagnostic system (CAD) leukocytes are separated from other blood cells. Besides, extraction of lymphocytes is done afterwards. CAD is proved to be quite useful in detecting leukemia based on texture and shape features. Then the extracted features are classified on support vector machine (SVM) classifier to reveal the occurrence of leukemic cells. CAD system provided the accuracy of 72.4 % for nucleus region, while 56.1 % for cytoplasm region. The combination of texture and shape features showed better results when compared to individuals since it provided the accuracy of 87.7 %.

Umpon [24] has segmented sand classified the white blood cell of bone marrow in microscopic images of bone marrow. He has presented the automated segmentation technique in the paper. The specific technique segments each blood cell intro three regions including background, cytoplasm, and nucleus. Segmentation performance of automated technique was evaluated while

comparing its performance with the manually segmented cells by the professional experts. The performance was compared by using probability of error (PE). The entire cell with nucleus was considered for segmenting image. Less than 10% probability of error (PE) was obtained after segmenting entire cell and nucleus.

Rajkumar et. al. [25] has discussed the renewed standard of diagnosing multiple myeloma. The members of international myeloma working group have added the attributes of CRAB features that include hypercalcaemia, renal failure, anemia and bone lesions besides other biomarkers. If the patients are identified with biomarkers along with above mentioned CRAB features, they are diagnosed with multiple myeloma who would otherwise be diagnosed with smouldering multiple myeloma. Since the delay in precise diagnosis and treatment of multiple myeloma could be harmful and injurious. In addition, the group provided the new standard that CRAB features should be included in the definition of multiple myeloma

Harada et. al. [26] has phenotypically distinguished normal plasma cells from mature myeloma cells via fluorescein isothiocyanate (FITC)-anti-CD38 antibody. Fluorescein isothiocyanate (FITC)-anti-CD38 antibody has the capacity to differentiate plasma or myeloma cells from other blood or hematopoietic cells present in bone marrow. Immunophenotypes of myeloma cells and normal plasma cells are examined phenotypically to differentiate normal plasma cells form myeloma cells. The analysis was done by applying two-colour flow cytometry with FITC-anti-CD38 antibody and phycoerythrin staining with various antibodies. Normal plasma cells from seven healthy contributors, four patients with chronic tonsillitis, one patient with idiopathic thrombocytopenic purpura, two patients with chronic lymphadenitis were considered as sample in the study. The authors concluded that normal plasma cells can be distinguished from malignant plasma cells via phenotypic analysis of plasma cells that have anti- CD56 and anti-CD19.

Riet et. al., [27] has studied the blood sample and bone marrow pf multiple myeloma patients to detect monoclonal B lymphocytes by rearrangement studies of immunoglobulin gene. Immunomagnetic beads were used to separate B lymphocytes. Ten patients comprising of newly diagnosed as well as previously treated patients were considered as a sample of this study. Monoclonal B lymphocyte was found in two out of ten patients. The findings showed that

presence of B lymphocytes in bone marrow are integral part of clone of myeloma. Further, these lymphocytes can circulate in the blood pf myeloma patients as well.

Ghnae et. al., [28] has segmented white blood cells or leukocytes from microscopic images by using watershed algorithm and K- means clustering. Since white blood cells identification is considered as a vital process to diagnose various diseases associated with blood. Initially, this identification is done by pathologists manually via optical microscope. The entire procedure is lengthy as it is time- consuming. Further, segmentation of white blood cells is fundamental step in order to establish a computer aided diagnostic system. The study was carried out to segment white blood cells from microscopic images. The authors have merged three techniques including k- means clustering, watershed algorithm and thresholding. The amalgamation of techniques was applied in three steps i.e., white blood cells were segmented from microscopic images, secondly, nuclei was extracted from the image of cell, and finally protruding nuclei and cells were separated. The technique was applied on 431 white blood cells from dataset. Then the results of segmentation were compared to manual segmentation done by the professional experts via three standards including similarity measures (T_S), sensitivity, and precision. The results have shown the similarity between the algorithm and manual segmentation by the experts. In addition, this proved that the proposed method could be quite fruitful for the pathologists in diagnosing various diseases

Sarrafzadeh, Omid, and Alireza Mehri Dehnavi [29] have used K- means clustering and region growing in order to segment cytoplasm and nucleus by using microscopic images. Their main aim is to detect leukocytes from microscopic image of a blood, then segment leukocytes into nucleus and cytoplasm. K means clustering is used to segment nuclei. Then, region growing is used isolate allied nuclei. Next, original input image is used to segment nucleus, while K- means clustering is used to segment cytoplasm. The results of the study show that segments of cytoplasm and nucleus proved to be useful for classification and feature extraction which will then use to discover leukemia.

Table 2.1.Comparison of Different Methods of Myeloma Cell Detection in Bone Marrow Aspiration

Authors (s), Year	Method	Dataset	Preprocessing	Segmentation	Feature Extraction	Classification	Performance %
▪ Saeedizadeh, Zahra, *et al.* 2016 [13]	Automatic recognition of myeloma cells in microscopic images using bottleneck algorithm, modified watershed and SVM classifier	50 images	Separation of images in to three components, median filter, contrast stretching	Cell Detection, Nuclei Splitting, Cell Splitting	Classification of normal plasma cells and myeloma cells	SVM (Support Vector Machine)	Sensitivity 96.52%, Specificity of 93.04%, Precision 95.28%
▪ Sarrafzadeh, Omid, et al. 2014 [14]	Detection of leukocytes from microscopic images of blood smear and	149 white blood cells	Manual determination of boundary of white blood cells	Fuzzy C-means clustering, morphological operations	Colour Features, Geometric Features, Texture	SVM (Support Vector Machine)	Accuracy 93%

	classification of leukocytes intro their types					Feature,		
▪ Mokhtar, N. R., et al. 2009 [15]	Contrast stretching techniques for acute leukemia images	Not mentioned	Local contrast stretching, global contrast stretching, partial contrast stretching	Image enhancement contrast stretching technique	Moment invariant Features, Veluchamy Features	SVM (Support Vector Machine)	Accuracy 93%
▪ Ongun, Guclu, et al. 2001 [16]	Extraction of features and classification of blood cells for automated DBC (differential blood count) system	Hematology laboratory of Hacettepe University Hospital, Ankara (108 bone marrow images of	Not mentioned	Active contour models/ snake- based segmentation	Shape based features, colour/ texture based features	k-NN (k-Nearest Neighbours), LVQ (Learning Vector Quantization), MLP (Multi-Layer Perceptron), SVM (Support Vector	Accuracy 87.98%

		258 white cells)				Machine)	
▪ Al-Quran, Samer Z., et al. 2007 [18]	Evaluation of bone marrow plasma cell percentage via microscope based differential count	University of Florida-Shands hospital (79 specimens of bone marrow of patients with multiple myeloma)	CD138 for assessing plasma cell numbers	Microscopic evaluation of plasma cells	Computer-assisted image analysis	Automated Cellular Imaging System (ACIS), Inc (currently Clarient, Inc, Aliso Viejo, CA)	Accuracy 95%
▪ Rajkumar, S. Vincent, et al. 2001 [19]	Techniques involved in estimation of bone marrow plasma cells of myeloma	Seventy-five patients	Not mentioned	Not mentioned	Not mentioned	Not mentioned	Highest estimate of Bone marrow Plasma Cell (BMPC) % is considered as suitable mean

							to identify patients with high involvement of bone marrow plasma cells
▪ Patel, Nimesh, and Ashutosh Mishra. 2015 [20]	Automated detection of leukemia using microscopic images	7 microscopic images	Histogram equalization, Zack algorithm	K- mean clustering	Colour features, geometric features, texture features, statistical features, Standard deviation, mean, area, perimeter	SVM (Support Vector Machine)	Accuracy 93.57%
▪ Cao, Haichao, Hong	Identification of bone marrow plasma cells	Five types of bone marrow	Colour transformation, Stepwise	K- mean clustering algorithm,	Size features, colour features,	SVM (Support Vector Machine), RF	Accuracy 87.49%

		cells (BMC)	Averaging Method (SAM)	Watershed algorithm	texture features, morphology features	(Random 0Forest), ANN (Artificial Neural Networks), Adaboost/ Enhanced learning, Bayesian Networks	
Liu, and Enmin Song. 2018 [21]							
▪ Brown, R. D., Luo, X. F., Gibson, J., Brisco, M., Sykes, P., Morley, A., &	Use of mRNA in situ hybridization to identify myeloma cells in the patients suffering from myeloma	7 patients	mRNA in situ hybridization (ISH)	DNA Sequencing	Not mentioned	Not mentioned	Not mentioned

|							
Joshua, D. 1998. [22]							
▪ Rawat, J., Singh, A., Bhadaur ia, H. S., & Virmani, J. 2015 [23]	Using CAD (computer-aided diagnosis) to identify leukemia via microscopic images	ALL-IDB1 (108 images), ALL-IDB2 (260 images)	Image enhancement	Various filtering operations	Shape based features, Gray level co–occurrence matrices (GLCM)	SVM (Support Vector Machine)	Accuracy 89.8%
▪ Theera-Umpon, N. 2005 [24]	Using segmentation technique to classify white blood cells (WBC) of microscopic bone marrow images	376 images of 6 types of white blood cells	Median filter	Fuzzy C-means (FCM) algorithm, mathematical morphology	Area of cell, shape of nucleus, size of nucleus, peak location of pattern spectrum, first and second granulometric	Neural network with five cross validation	Less than 10% probability of error (PE) was obtained after segmenting entire cell and nucleus

					moments of nucleus		
▪ Ghane, Narjes, et al. 2017 [28]	Using microscopic images to segment cells and nuclei of white blood cells (WBCs) through K-Means Clustering and Modified Watershed Algorithm	45 images from 431 white blood cells of six patients	Minimum Filter, Colour adjustment method	K-Means Clustering, Modified Watershed Algorithm, Otsu thresholding method, Gradient method	Not mentioned	Not mentioned	Nucleus segmentation (similarity measures 92.07%, precision 96.07%, sensitivity 94.30%), cell segmentation (similarity measures 92.93%, precision 97.41%, sensitivity 93.78%)
▪ Sarrafza deh,	Microscopic imag based	Not mentioned	Median Filter	K-means clustering,	Colour features	Not mentioned	Proposed method is ideal

Omid, and Alireza Mehri Dehnavi. 2015 [29]	segmentation using K- means clustering and region growing for automated detection of leukemia			region growing			to determine leukocytes

Proposed Methodology

3.1   Dataset

The data set of current study consists of bone marrow aspiration slides. The slides were prepared after fitting camera with microscope, so right after viewing cells of data bases, a high-resolution digital camera Nikon1 V1 coupled to Nikon Eclipse 50i light microscope was accommodated to store the seen images. The contribution by distinct professionals leads to the reliability of data set when compared to commonly available data set. Altogether, the dataset was provided by the pathology department of Alzahra hospital, Isfahan, Iran.
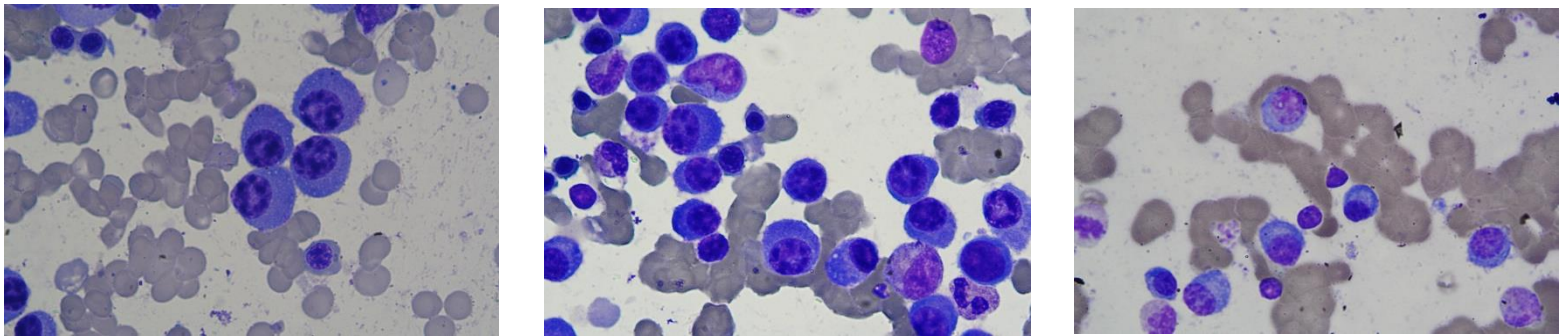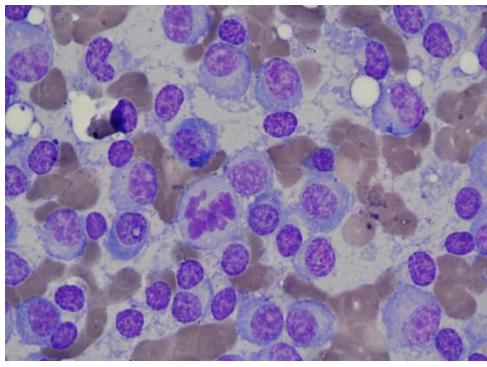

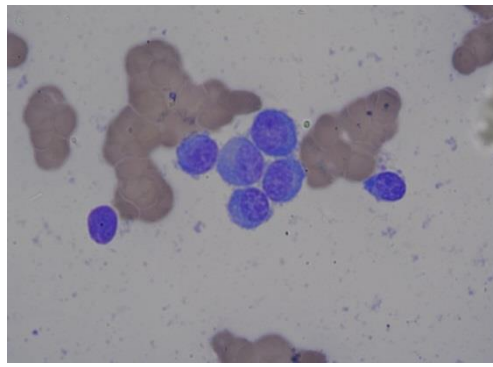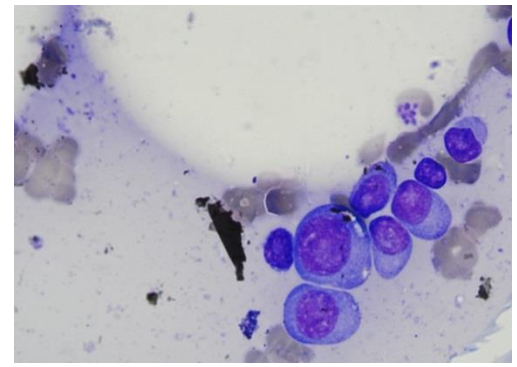**(A)**   **(B)**   **(C)**

Figure 3.1: (A), (B), and (C): Type of plasma cells- Normal plasma cells

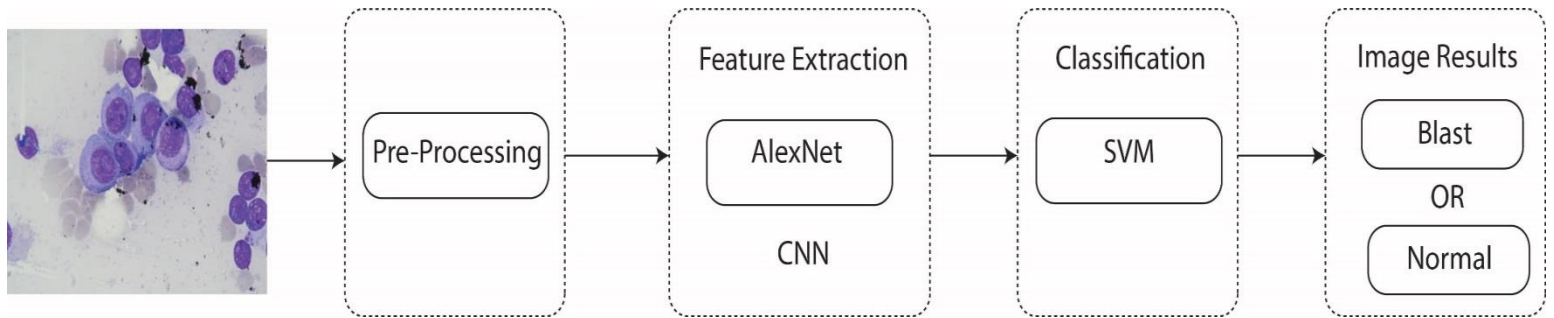**Figure (D), (E), and (F): Type of plasma cells- Blast plasma cells**

Figure 3.2: Flowchart of Proposed Methodology

Proposed methodology is followed by preprocessing that is subsequently followed by three stages of enhancement, including image separation into components, application of median filter, and the linear contrast stretching. Afterwards, CNN is followed by AlexNet. Finally, SVM classifier is discussed.

3.2 Preprocessing

An exterior blood image contains erythrocytes, leucocytes, thrombocytes and background as well. They all could contain spot noise. Henceforth, preprocessing procedures are used to refine the images. This refinement helps in overcoming the probability of high variation in contrast. Further, the probability of high variation is due to diversity of alterations in staining and light. Altogether, preprocessing is carried out to apply filter and contrast enhancement process [13].

Enhancement procedure comprises of three steps:

1. Separation of microscopic image into three components which are R, G, and B

2. Application of median filter on R, G, and B components (to lessen the spot noise and sustaining the edges to the full).

Median filtering is a nonlinear operational technique applied on images to lessen "salt and pepper" noise.

3. Linear contrast stretching (Mokhtar *et al.*, 2009) is applied on input image via following formula:

$$Ip\ (x,y)\ =\ \frac{55\ \times\ [I0\ (x,y)\ -\ m]}{M\ -\ m}$$

➤ (*x, y*) stands for pixel

➤ I_p (*x, y*) is representing grey level of output pixel

➤ I_0 (*x, y*) is representing grey level of input pixel

➤ *M* stands for upper limit of contrast stretching

➤ *m* stands for lower limit of contrast stretching

The above mentioned method proved to be very useful in achieving improved results of segmentation since the contrast between cytoplasm of erythrocytes and leucocytes is intensified. There are definite and distinct upper as well as lower limits for three channels of filtered R, G, and B images. Moreover, CAT function is applied to merge R, G, and B components. Then after saving the image, the consequences of median filter application as well as contrast stretching on R, G, and B images are illustrated in figure 3.3. The colour of cytoplasm, erythrocytes, and nucleus is altered. Moreover, leucocytes could be easily and efficiently identified from the concerned image.
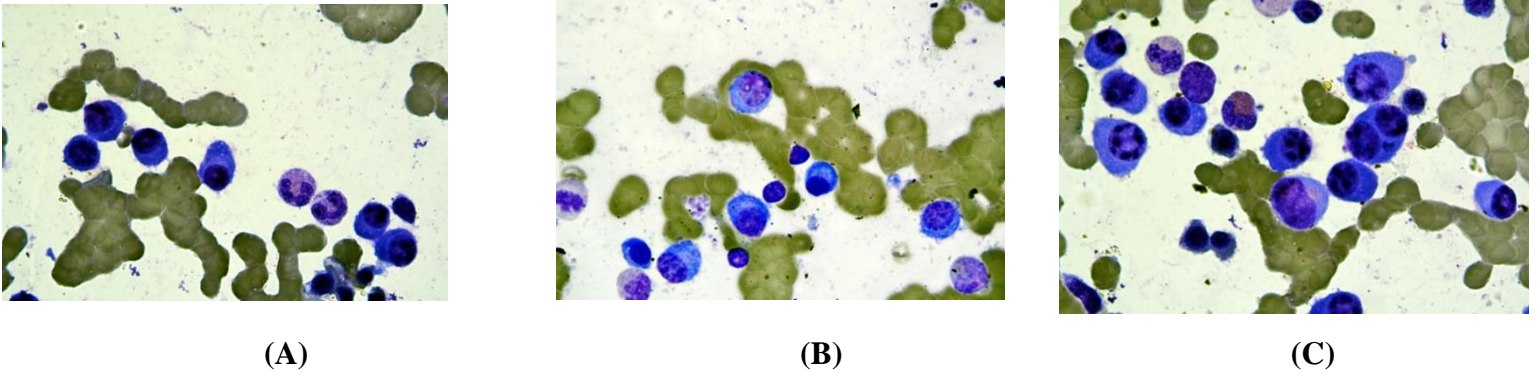
<center>(A)   (B)   (C)</center>

Figure 3.3: (A), (B), and (C): Type of plasma cells- Normal plasma cells
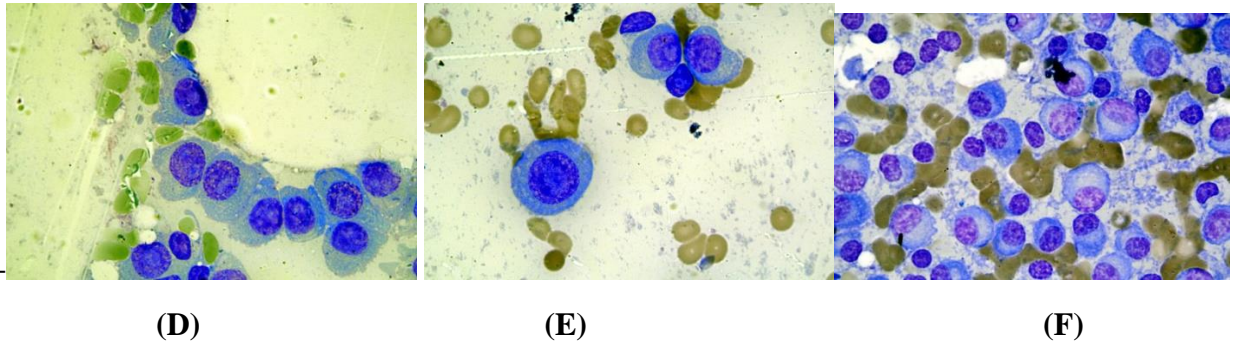
<center>(D)   (E)   (F)</center>

Figure (D), (E), and (F): Type of plasma cells- Blast plasma cells

In short, this preprocessing technique is useful in distinguishing between erythrocytes and cytoplasm. Since there is too much homogeneity in brightness and color of erythrocytes and cytoplasm of the genuine image, so enhancement algorithm is employed to distinguish between them.

3.3 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a network that consists of one or sometimes more than one convolutional layers. Additionally, convolution layers combine along with subsampling step. Convolutional layers have the ability to amend the data representation via various filters.

CNN is subsequently followed by thoroughly connected layers as in a standard multilayer neural network. CNN is used to view the two-dimensional structure of an input image. Recognition of object and classification of image are done through CNN. One of the noticeable advantage of

CNNs is that they can be easily trained. CNN has lesser parameters than fully connected networks which have similar amount of hidden units.
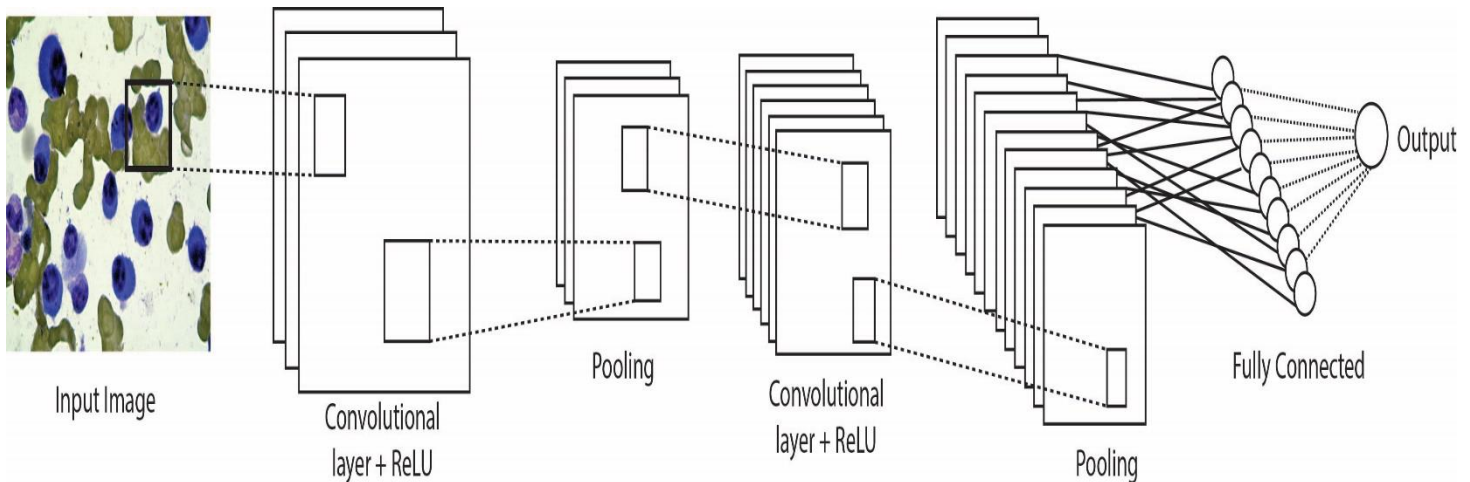

Figure 3.4: Simplified Illustration of the CNN Architecture

The accommodation of CNN in extraction has been widely used now a days amongst other descriptors presented in the state of-the-art [30]. The excessive set of data is represented via exclusive features in feature extraction procedure. There are numerous features that can be extracted from an input image. Colour, texture, and shape to name a few.

An activation function is used after convolutional layers. Besides, non-linear transformations in the data are performed by activation functions. The purpose is to induce outputs that are linear as well as separable. Rectified Linear units (ReLu) is the most common objective accommodated in this principle. ReLu is mentioned in the following equation.

f(x) = max(0,x)

> ➢ 'x' is representing the input to a neuron.

CNN comprising of convolutional and subsampling sub-layers is illustrated in the following diagram. Previous layers are overcome by the pooling layer. Henceforth, pooling layer reduces

the quantity of features of resultant data. The pooling layer succeeds the previous layers and reduces the amount of features of the resulting data. Features of descriptors are assembled via layers which are entirely connected, since their function is to assemble all the features of descriptors, in order to classify the final layer.

The above figure illustrates the constitution of various layers discussed earlier. The contemporary authors have discussed two distinct means of exercising the power of a CNN. At the very basic level, CNN could implement the training via data. However, there should be a large set of data. Secondly, learning is transmitted by employing already trained networks. In the current study, learning transmittal technique is deployed. Moreover, CNN is skilled by applying large set of images. Generic features are incorporated by CNN training. This training enables the facilitation in the pertinence of CNN in trivial databases. This proficiency allows to employ various tasks such as in feature extraction from objects, diseases, and face images. The accuracy of consequences is dependent on likeliness of images with the base adopted in feature extraction and the trained images.

Castellucio et al. have discussed two options of learning transfer or transmittal mentioned above. First is the proper tuning of network which includes the modification of structure that incorporates the freezing of various layers presented in the high level. Second option is extracting last layer of network that is entirely connected, and achieved via input image.

3.4   AlexNet

AlexNet is the architecture of CNN. Krizhevsky et al. have developed AlexNet for the competition of ILSVRC in 2010. The completion was carried out to classify and train the database of ImageNet. AlexNet is constituted of eight layers that should be skilled enough, along with five convolutional layers and eight fully connected layers from FC6-8. Besides, there are features in FC6-7. However, in last eighth layer, there are total number of categories that should be classified. Convolutional layers should have filters of 5x5 and 7x7 succeeded by three fully connected layers and max- pooling layers [30]. Besides, there are 4096 features extracted from the fully connected layer of each individual architecture of CNN. After extraction process, there is the fusion of attributes from each of the architecture as a consequence of resultant vector in thoroughly connected network. After the fusion, there are 12288 characteristics.
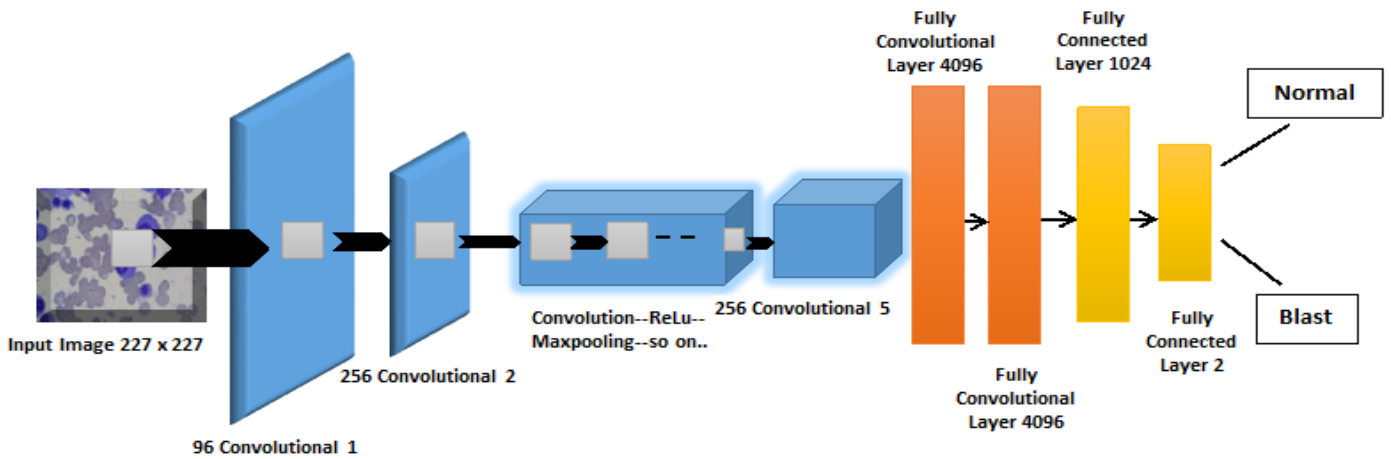
Figure 3.5: AlexNet Architecture

3.5   SVM

For classifying myeloma cell, there is a need of classifier after determining features. Moreover, the classifier should be powerful enough to provide fine results. For classifying, SVM is used since SVM is one of the methods widely used for classification [30]. This methods was primarily designed for two class problem (Gunn, 1998). The veracity of SVM is determined individually for the features or characteristics of nine colour sub- bands i.e. (R, G, B), (H, S, V) and (L, a, b). Entropy of nucleus is the outstanding feature for B sub- band from RGB colour system, whereas total entropy is the finest feature for V sub- band from HSV colour system, because entropy of nucleus and total entropy are most known for determining various feature including colour sub- bands. The particular features are determined by using the following equation:

Entropy: $\dfrac{\sum_{X=1}^{M} \sum_{Y=1}^{N} I_i\,(x,y)\left(-\ln I_i(x,y)\right)}{M \times N}$

➤ 'x' stands for input vector

SVM has the ability of mapping out the input vector into high dimensional feature space. This is done by choosing a non- linear mapping which include polynomials, sigmoid functions and radial basis functions (Gunn, 1998; Smola & Scholkopf, 2004). Sigma =3 with radial basis function is used in the current study. K-fold cross-validation with K equal to 10 to assess the consequences of classification. K-1 old is used as a trained set however at last fold is used as an evaluated set in cross validation of K- fold. The above mentioned procedure is repeated 10 times which is the supposed value of "K". This repetition leaves a particular fold every time for evaluating purpose.

3.6. Multiple Myeloma Detection Using Transfer Learning

Multiple myeloma is detected via pre- trained AlexNet in the current study. A software MATLAB is employed to run the architecture in order to provide an effective environment for neural networks and deep learning execution. Input images were RGB which have resolution of 227 x 227 pixels. AlexNet comprised of 5 convolutional layers along with 3 max polling layers. Besides, every convolutional layer in AlexNet architecture has ReLU (rectifed linear unit) with it. Last three layers of the pre trained AlexNet were removed for transfer learning. Those layers were fully connected trained for 1000 categories classification, softmax layer and classification layer. The architecture was tuned for two reasons after removing the layers, i.e. detection of MM and its classification into normal and blasted. So another fully connected layer is added after this.

The rate of change of network is usually handled by learning rate. In this network we didn't change the learn rate of previous layers which were before the last 3 layers, instead we increased the learn rate of our newly added layers to make them update quickly than the original layers, so that we can learn the new weight of these layers faster. In this study batch size was set to 28 and max epochs to 100. Training was carried out on a laptop having Intel (R) core(TM)i3-5010u,x64 based processor which took around 40 minutes.

CHAPTER # 4

Result and Analysis

Microscopic bone marrow images are utilized to apply the proposed methodology. There are fifty images of data set altogether on which algorithm is applied. The images comprise of thirty-four blasted plasma cell (BPC), and 16 normal plasma cells. The horizontal and vertical resolution of the images of data set is 300 dpi and 300 dpi respectively. The ratio of images of data set is 60:40, since 60 % of images are considered as trained data set, while 40% are taken as test data set. Besides, images of data set contain normal plasma cells, myeloma cells and other types of cells as well. The images were collected at different times from various bone marrow smears contributed by various professionals.

The attainment of the algorithm in detecting myeloma cells is evaluated via two specifically defined measures. SS is representing "sensitivity" that is the probability detected by the cell if it has been identified as plasma cell by the professionals. SC is representing "specificity" that is the probability detected by the cell if it has been identified as marrow cell by the professionals. Besides, ACC stands for "accuracy' that is the probability in case there is true detection of cells.

SS, SC and ACC are calculated as follow for SVM:

Table 4.1: T_P, T_N, F_P, F_N Values for SVM

	+	-
+	T_P 1	F_P 0
-	F_N 0	T_N 1

$$SS= \frac{TP}{TP+FN}$$

$$= \frac{1}{1+0}$$

$$= 1$$

$$= 100\%$$

$$SC = \frac{TN}{TN+FP}$$

$$= \frac{1}{1+0}$$

$$= 1$$

$$= 100\%$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{1+1}{1+1+0+0}$$

$$= \frac{2}{2}$$

$$= 1$$

$$= 100\%$$

> ➢ TP stands for "true positive" that is the total number of plasma cells which are identified as myeloma cells by the professionals as well as algorithm
> ➢ TN stands for "true negative" that is the total number of plasma cells which are identified as normal plasma cells by the professionals as well as algorithm
> ➢ FP stands for "false positive" that is the total number of plasma cells which are identified as normal plasma cells by the professionals, and myeloma cells by the algorithm
> ➢ FN stands for "false negative" that is the total number of plasma cells which are identified as myeloma cells by the professionals, and normal plasma cells by the algorithm

SS, SC and ACC are calculated as follow for KNN:

Table 4.2: T_P, T_N, F_P, F_N Values for KNN

	+	-
+	T_P 0.9286	F_P 0.0714
-	F_N 0	T_N 1

SS= $\dfrac{TP}{TP+FN}$

= $\dfrac{0.9286}{0.9286+0}$

= 1

= 100%

SC= $\dfrac{TN}{TN+FP}$

= $\dfrac{1}{1+0.0714}$

= $\dfrac{1}{1.0714}$

= 0.93

= 93%

ACC= $\dfrac{TP+TN}{TP+TN+FP+FN}$

= $\dfrac{0.9286+1}{0.9286+1+0.0714+0}$

= $\dfrac{1.9286}{2}$

= 0.96

$$= 96\ \%$$

SS, SC and ACC are calculated as follow for D-tree:

Table 4.3: T_P, T_N, F_P, F_N Values for D-tree

	+	-
+	T_P 0.7143	F_P 0.2857
-	F_N 0	T_N 1

$$SS = \frac{TP}{TP+FN}$$

$$= \frac{0.7143}{0.7143+0}$$

$$= 1$$

$$= 100\%$$

$$SC = \frac{TN}{TN+FP}$$

$$= \frac{1}{1+0.2857}$$

$$= \frac{1}{1.2857}$$

$$= 0.77$$

$$= 77\ \%$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{0.7143+1}{0.7143+1+0.2857+0}$$

$$= \frac{1.7143}{2}$$

$=$ 0.85

$=$ 85 %

Table 4.4: Comparison of Performance of SVM, KNN, D- Tree and CNN for Dataset without Preprocessing

	Sensitivity %	Specificity %	Accuracy %
SVM	89	71	75
KNN	96	84	80
D- Tree	79	73	70
CNN	100	92	95

Table 4.4: Comparison of Performance of SVM, KNN, D- Tree and CNN for Preprocessed Dataset

	Sensitivity %	Specificity %	Accuracy %
SVM	100	100	100
KNN	100	93	96
D- Tree	100	87	85
CNN	100	100	100

The fine- tuned Alexnet was employed to automatically detect multiple myeloma from the dataset. Dataset A contains the R,G, and B images. Datset B contains H,S, and V images. Y,

CB, and CR images are present in dataset C, and dataset D contains the H, CB, and CR images in in it. However, if none of the classifiers is applied to classify the images into normal and blast cells, and merely convolutional neural network is applied, then the accuracy is 100%.

Table 4.5: Classification Accuracies (%) for Different Colour Combination Datasets

	Training Samples	Dataset A (R,G,B)	Dataset B (H,S,V)	Dataset C (Y, CB, CR)	Dataset D (H, CB, CR)
MM Detection	50	100	80	75	65

CHAPTER # 5

Conclusion and Future Work

5.1   Conclusion

In the current study, myeloma cells in bone marrow images are detected automatically. Various steps including preprocessing, feature extraction and segmentation are used to detect myeloma cells in bone marrow because microscopic examination is not sufficient enough. This inadequacy is because of involvement of human factors i.e., fatigue, stress, proficiency, experience etc. Also, these factors could affect the outcome. Therefore, there is utmost need of some automated method to detect myeloma cells in bone marrow aspirations to achieve accuracy in performance which would efficiently assist the professionals to detect multiple myeloma.

Convolutional neural network is used to detect multiple myeloma through microscopic images of blood. Moreover, pre- trained AlexNet was arrayed instead of developing an architecture from scratch. Input images are classified into normal and blast by substituting last layer of pre- trained AlexNet by two new layers. Besides, there is the need of classifier in order to classify myeloma cells. Hence, SVM is used to classify myeloma cells into normal and blast.

Purposed methodology was compared with the previous literature in terms of total number of images in the data set and their performance as well. Our methodology unlike previous ones performed well in terms of carrying out 100 % accuracy without employing any microscopic segmentation.

5.2   Future Work

In the current study, myeloma cells are detected automatically by using CNN along with support vector machine classifier. In future, we can employ large dataset as compared to the current study. Also, CNN could be implemented from scratch. The neural network could be trained from the very beginning. Future researcher can employ diverse network for detection and classification of multiple myeloma. This would help the hematologists and pathologists to detect the myeloma cells in a much better way in everyday life.

References

1. Guyton, A. & Hall, J. (2006) *Textbook of Medical Physiology*. 11th edn. Elsevier Saunders, Philadelphia.

2. SciCom Zambia website http://web.mit.edu/scicom/www/blood.html

3. Theml, H., Diem, H. & Haferlach, T. (2004) *Color Atlas of Hematology: Practical Microscopic and Clinical Diagnosis*, Thieme, Stuttgart, New York.

4. Rajkumar SV: Multiple myeloma: 2013 update on diagnosis, risk-stratification, and management. Am J Hematol 2013; 88:226- 235.

5. Multiple Myeloma Research Foundation (MMRF) website https://www.themmrf.org/multiple-myeloma/

6. Bataille R, Harousseau JL. Multiple myeloma. N Engl J Med. 1997; 336:1657-1664.

7. Lahtinen, R., Lakso, M., Palva, I., Virkkunen, P. & Elomma, I. (1992). Randomised, placebo controlled multicenter trial of clodronate in multiple myeloma. *Lancet*, 340, 1049- 1052.

8. Fleisch, H. (1991) Bisphosphonates: pharmacology and use in the treatment of tumour-induced hypercalcaemic and metastatic bone disease. *Drugs,* 42, 919- 944.

9. McCloskey, E.V., MacLennan, I.C.M., Drayson, M., Chapman, C., Dunn, J. & Kanis, J.A. (1996) Effect of clodromnate on morbidity and mortality in myelomatosis. (Abstarct). *Annals of Oncology,* 7, (Suppl. 5), Abstr. 4120.

10. Minges Wols, H.A. (2006) Plasma cells. In: eLS, editor. USA: John Wiley & Sons Ltd, Chichester. 1038, 1–810.

11. Multiple Myeloma Research Foundation (MMRF) website https://www.themmrf.org/multiple-myeloma/diagnosis/myeloma-diagnostic-testing/

12. Smith, A, *et al.* Guidelines on the diagnosis and management of multiple myeloma. *British Journal of Haematology, 132* (2005)*,* 410–451

13. Saeedizadeh, Zahra, *et al*. "Automatic recognition of myeloma cells in microscopic images using bottleneck algorithm, modified watershed and SVM classifier." *Journal of microscopy* 261.1 (2016): 46-56

14. Sarrafzadeh, Omid, et al. "Selection of the best features for leukocytes classification in blood smear microscopic images." *Medical Imaging 2014: Digital Pathology*. Vol. 9041. International Society for Optics and Photonics, 2014.

15. Mokhtar, N. R., et al. "Image enhancement techniques using local, global, bright, dark and partial contrast stretching for acute leukemia images." *Proceedings of the world Congress on Engineering*. Vol. 1. 2009.

16. Ongun, Guclu, et al. "Feature extraction and classification of blood cells for an automated differential blood count system." *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*. Vol. 4. IEEE, 2001.

17. Shipman, Claire M., et al. "Bisphosphonates induce apoptosis in human myeloma cell lines: a novel anti-tumour activity." *British journal of haematology* 98.3 (1997): 665-672.

18. Al-Quran, Samer Z., et al. "Assessment of bone marrow plasma cell infiltrates in multiple myeloma: the added value of CD138 immunohistochemistry." *Human pathology* 38.12 (2007): 1779-1787.

19. Rajkumar, S. Vincent, et al. "Methods for estimation of bone marrow plasma cell involvement in myeloma: predictive value for response and survival in patients undergoing autologous stem cell transplantation." *American journal of hematology* 68.4 (2001): 269-275.

20. Patel, Nimesh, and Ashutosh Mishra. "Automated leukaemia detection using microscopic images." *Procedia Computer Science* 58 (2015): 635-642.

21. Cao, Haichao, Hong Liu, and Enmin Song. "Bone marrow cells detection: A technique for the microscopic image analysis." *arXiv preprint arXiv:1805.02058* (2018).

22. Brown, R. D., Luo, X. F., Gibson, J., Brisco, M., Sykes, P., Morley, A., & Joshua, D. (1998). Longitudinal analysis of circulating myeloma cells detected by allele specific mRNA in situ hybridization. *American journal of hematology*, *58*(4), 273-277.

23. Rawat, J., Singh, A., Bhadauria, H. S., & Virmani, J. (2015). Computer aided diagnostic system for detection of leukemia using microscopic images. *Procedia Computer Science*, *70*, 748-756.

24. Theera-Umpon, N. (2005, August). White blood cell segmentation and classification in microscopic bone marrow images. In *International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 787-796). Springer, Berlin, Heidelberg.

25. Rajkumar, S. Vincent, et al. "International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma." *The lancet oncology* 15.12 (2014): e538-e548.

26. Harada, Hironori, et al. "Phenotypic difference of normal plasma cells from mature myeloma cells." *Blood* 81.10 (1993): 2658-2663.

27. Van Riet, Ivan, et al. "Detection of monoclonal B lymphocytes in bone marrow and peripheral blood of multiple myeloma patients by immunoglobulin gene rearrangement studies." *British Journal of Haematology* 73.3 (1989): 289-295.

28. Ghane, Narjes, et al. "Segmentation of white blood cells from microscopic images using a novel combination of K-means clustering and modified watershed algorithm." *Journal of medical signals and sensors* 7.2 (2017): 92. Assessed from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437768/

29. Sarrafzadeh, Omid, and Alireza Mehri Dehnavi. "Nucleus and cytoplasm segmentation in microscopic images using K-means clustering and region growing." *Advanced biomedical research* 4 (2015) Assessed from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617159/

30. Vogado, Luis Henrique Silva, et al. "Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-trained Convolutional Neural Networks." *Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference on*. IEEE, 2017.