IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 22-29
www.iosrjournals.org

An Overview on Acute Lymphocytic Leukemia Detection using Cell Image Segmentation

Ansha Beevi S.[1], Remya R.S.[2]
[1] *(Department of Computer Science, College of Engineering, Karunagappally, India)*
[2] *(Department of Computer Science, College of Engineering, Karunagappally, India)*

ABSTRACT: *Leukemia is a disease that affects blood forming cells in the body. Early detection of the disease is necessary for proper treatment management. Abnormal white blood cells or blasts play important role for hematologists in their diagnostic process. The diagnosis and treatment of diseases have been significantly simplified by the Differential Counting of white blood cells as it furnishes critical information required by pathologists. The practice of manual counting of the white blood cells suffer from the obvious disadvantages associated with human errors. Naturally an automated process would be the ideal solution to such a problem. Digital image processing technique could help them by enhancing the visibility in their analysis and diagnosis. White blood cell segmentations are an important research issue in Hematology. Our research proposes a segmentation of nucleus and cytoplasm of white blood cell slides. In this paper, an overview of various techniques used for Leukemia Detection using cell image segmentation is discussed.*

Index terms: *Bayes classification, geometric features, GLCM, watershed segmentation, white blood cell.*

I. INTRODUCTION

The body is made up of trillions of living cells. Normal body cells grow, divide into new cells, and die in an orderly way. During the early years of a person's life, normal cells divide faster to allow the person to grow. After the person becomes an adult, most cells divide only to replace worn-out or dying cells or to repair injuries. Cancer begins when cells in a part of the body start to grow out of control. There are many kinds of cancer, but they all start because of out-of-control growth of abnormal cells. Cancer cell growth is different from normal cell growth. Instead of dying, cancer cells continue to grow and form new, abnormal cells. Cancer cells can also invade (grow into) other tissues, something that normal cells cannot do. Growing out of control and invading other tissues are what makes a cell a cancer cell.

Cells become cancer cells because of damage to DNA. DNA is in every cell and directs all its actions. In a normal cell, when DNA gets damaged the cell either repairs the damage or the cell dies. In cancer cells, the damaged DNA is not repaired, but the cell doesn't die like it should. Instead, this cell goes on making new cells that the body does not need. These new cells will all have the same damaged DNA as the first cell does. People can inherit damaged DNA, but most DNA damage is caused by mistakes that happen while the normal cell is reproducing or by something in our environment. Sometimes the cause of the DNA damage is something obvious, like cigarette smoking. But often no clear cause is found. In most cases the cancer cells form a tumor. Some cancers, like leukemia, rarely form tumors. Instead, these cancer cells involve the blood and blood-forming organs and circulate through other tissues where they grow.

Different types of cancer can behave very differently. For example, lung cancer and breast cancer are very different diseases. They grow at different rates and respond to different treatments. That is why people with cancer need treatment that is aimed at their particular kind of cancer. Not all tumors are cancerous. Tumors that aren't cancer are called benign. Benign tumors can cause problems they can grow very large and press on healthy organs and tissues. But they cannot grow into (invade) other tissues. Because they can't invade, they also can't spread to other parts of the body (metastasize). These tumors are almost never life threatening.

1.1 Leukemia

Leukemia is a disease that affects blood forming cells in the body. Early detection of the disease is necessary for proper treatment management. An estimated 48,610 new cases of leukemia are expected in 2013. Leukemia is a cancer of the bone marrow and blood and is classified into four main groups according to cell type and rate of growth: acute lymphocytic (ALL), chronic lymphocytic (CLL), acute myeloid (AML), and chronic myeloid (CML). Almost 90% of leukemia cases are diagnosed in adults 20 years of age and older, among whom the most common types are CLL (38%) and AML (30%). Among children and teens, ALL is most

common, accounting for 75% of leukemia cases (see Childhood Cancer, page 11). From 2005 to 2009, overall leukemia incidence rates increased slightly by 0:4% per year.

An estimated 23,720 deaths are expected to occur in 2013. Death rates for leukemia have been declining for the past several decades; from 2005 to 2009, rates decreased by 0:8% per year among males and by 1:4% per year among females.

1.2 Signs and symptoms

Symptoms may include fatigue, paleness, weight loss, repeated infections, fever, bruising easily, and nosebleeds or other hemorrhages. In acute leukemia, these signs can appear suddenly. Chronic leukemia typically progresses slowly with few symptoms and is often diagnosed during routine blood tests. Patients with CLL may experience swollen lymph nodes or pain in the upper left abdomen due to an enlarged spleen.

1.3 Risk factors

Exposure to ionizing radiation increases risk of several types of leukemia (excluding CLL). Medical radiation, such as that used in cancer treatment, is a substantial source of radiation exposure. Leukemia may also occur as a side effect of chemotherapy. Children with Down syndrome and certain other genetic abnormalities are at increased risk of leukemia. Workers in the rubber-manufacturing industry also have an increased risk. Recent studies suggest that obesity increases risk of leukemia.

Some factors are most closely associated with specific types of leukemia. Family history is one of the strongest risk factors for CLL. Cigarette smoking is a risk factor for AML, and there is limited evidence that parental smoking and maternal exposure to paint increases the risk of childhood leukemia. Exposure to certain chemicals, such as formaldehyde and benzene (a component in cigarette smoke and gasoline that has become more regulated due to its carcinogenicity), also increases risk of AML. Infection with human T-cell leukemia virus type I (HTLV-I) can cause a rare type of leukemia called adult T-cell leukemia/lymphoma. The prevalence of HTLV-I infection is geographically localized and is most common in southern Japan and the Caribbean; infected individuals in the US tend to be descendants or immigrants from endemic regions.

1.4 Early detection

Leukemia can be difficult to diagnose early because symptoms often resemble those of other, less serious conditions. When a physician does suspect leukemia, diagnosis can be made using blood tests and a bone marrow biopsy.

1.5 Treatment

Chemotherapy is the most effective method of treating leukemia. Various anticancer drugs are used, either in combination or as single agents. Imatinib (Gleevec), nilotinib (Tasigna), and dasatinib (Sprycel) are very effective drugs that are targeted at the genetic abnormality that is the hallmark of CML. Imatinib and dasatinib are also FDA-approved to treat a type of ALL with the same genetic defect. People diagnosed with CLL that is not progressing or causing symptoms may not require treatment. Recent clinical trials have shown those adults with AML who are treated with twice the conventional dose of daunorubicin experience higher and more rapid rates of remission. Antibiotics and transfusions of blood components are used as supportive treatments. Under appropriate conditions, stem cell transplantation may be useful in treating certain types of leukemia.

1.6 Survival

Survival rates vary substantially by leukemia type, ranging from a 5-year relative survival of 25% for patients diagnosed with AML to 82% for those with CLL. Advances in treatment have resulted in a dramatic improvement in survival over the past three decades for most types of leukemia. For example, from 1975-1977 to 2002-2008, the 5-year relative survival rate for ALL increased from 41% to 68% overall, and from 58% to 91% among children. In large part due to the discovery of targeted cancer drugs like Imatinib, the 5-year survival rate for CML increased from 31% for cases diagnosed during 1990-1992 to 56% for those diagnosed during 2002-2008.

Acute Lymphocytic Leukemia (ALL), also known as acute lymphoblastic leukemia is a cancer of the white blood cells, characterized by the overproduction and continuous multiplication of malignant and immature white blood cells (referred to as lymphoblast or blasts) in the bone marrow. It is fatal if left untreated due to its rapid spread into the bloodstream and other vital organs.

ALL produces a lack of healthy blood cells due to an abnormal number of malignant and immature white blood cells. It mainly affects young children and adults over 50. Early diagnosis of the disease is fundamental for the recovery of patients especially in the case of children.

Unfortunately, the initial symptoms of ALL are quite not specific: generalized weakness, anemia, frequent fever and infections, weight loss and/or loss of appetite, excessive bruising or bleeding from wounds, nosebleeds, bone pain, joint pains, breathlessness, enlarged lymph nodes, liver and/or spleen. If the described symptoms are present, blood tests such as a full blood count, renal function, electrolytes and liver enzymes and blood count have to be done. Clinical suspicion alone may be the only reason to perform a bone marrow biopsy, which is the next step in the diagnostic process. Bone marrow is examined for blasts, cell counts and other signs of disease. Pathological examination, cytogenetic and immunophenotyping are common diagnostic analysis. Once the blast-cells invasion starts, blast cells can be detected in peripheral blood.

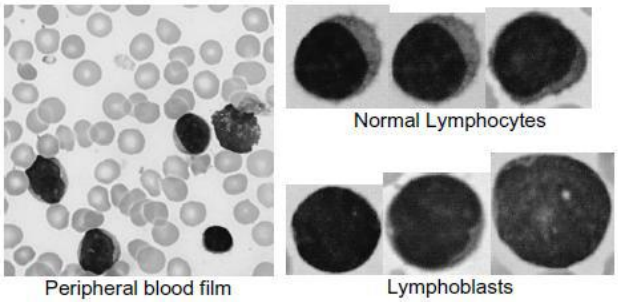

Fig.1: Peripheral blood film (left) with white cells marked with colorant. On the right, examples of normal (top) and blast (down) lymphocytes acquired with 1000x microscope.

Figure 1 shows the microscope image of a blood film (left) and it plots three examples of normal lymphocytes and three lymphoblastic cells (right). Images have been digitalized by the optical microscope by using a CCD and then acquired by a frame-grabber system. Principal cells present in the blood are the red blood cells, and the white cells (leucocytes). Leukocyte cells containing granules are called granulocytes (composed by neutrophil, basophil, and eosinophil). Cells without granules are called agranulocytes (lymphocyte and monocyte). ALL symptoms are associated only to the lymphocytes.

Hence, the observation of the peripheral blood _lm by expert operators is one of the diagnostic procedures available to evaluate the presence of the acute leukemia. This analysis suffers from slowness and also presents a not standardized accuracy since it depends on the operator's capabilities and tiredness. Conversely, the morphological analysis just requires an image –not a blood sample- and hence is suitable for low-cost, standard-accurate, and remote diagnostic systems.

In today's diagnostic paradigm, microscopic imaging technology has immense contributions in generating fruitful medical images, which essentially become the basis for medical experts to make better decisions. In practice, experts like radiologists and pathologists, visualize the abnormalities if any, in the images through microscope based on their subjective knowledge from the view point of intensity, morphology, texture etc. based features. Usually small scale differences in the features are overlooked by human eyes especially for the border-line diagnostic scenario. It is more worthwhile to develop computer-assisted automated screening scheme for automatically characterizing the abnormalities, especially in complicated cases where experts fail to take decision. In doing this, microscopic information needs to be analyzed quantitatively maintaining the biological integrity in the system.

Leukemia is a group of blood diseases affecting the blood cells and most commonly white blood cells. Leukemia is characterized by overproduction of abnormal (or immature) white cells which are unable to fight infection. There are two main types of acute leukemia: Acute Lymphoblastic Leukemia (ALL) and Acute Myelogenous Leukemia (AML). Acute leukemia is a rapidly progressing disease that affects mostly cells that are unformed (not yet fully developed or differentiated). ALL is the most common type of leukemia in young children. This disease also affects adults, especially those aged 65 and older.

In order to classify the abnormal cells in their particular types and subtype of leukemia, an expert operator will observed some cells under a light microscopy looking for the abnormalities presented in the nucleus or cytoplasm of the cells. This classification is very important to determine which treatment should be given to the patient [4]. However, this analysis suffers from slowness and it presents a not standardized accuracy since it depends on the operator's capabilities and tiredness.

Regardless of advanced techniques i.e. flow cytometer, immune phenotyping, molecular probing etc, microscopic examination of blood slides still remains as a standard leukemia diagnosis technique. Hence microscopic examination is the most economical way for initial screening of leukemia patients. Manual examination of the slides are subjected to bias i.e. operator experience, tiredness etc resulting with inconsistent and subjective reports. So there is always a need for a cost effective and robust automated system for leukemia screening which can greatly improve the output without being influenced by operator fatigue.

II. OVERVIEW

The various steps involved in Leukemia detection are pre-processing, segmentation of white blood cells, feature extraction and then classification.

2.1. Pre-processing

During image acquisition, images are saved in JPEG format to use less computational requirements and to ensure not to have over-segmentation in watershed algorithm as image will have more details. As it is commonly the case, acquired images have all blood elements colors close to background color, red blood cells are clustered with white blood cells and the presence of noise and stain in the blood slides is significant (Hengen et al., 2002). To overcome or reduce the effect of such factors, Juma Al-Muhairy et al. [2] posteriori standardized the image by increasing their contrast.

During image acquisition and excessive staining, the images will be disturbed by noise. The noise may be due to illumination or shadows that make region of interest (ROI) appear as blurred image region. Background will be excluded since ROI will be white blood cells. Image enhancement was done as the contrast enhancement technique [11], is capable to improve the medical image quality (figure 2).

In [8], the input image is filtered with Wiener filter to remove the background noise which is directly implanted within the image during image acquisition. A Wiener filter (a type of linear filter) tailors itself to the local image variance. Where the variance is large, wiener performs little smoothing. Where the variance is small, wiener performs more smoothing. For enhancement of the noise free image Laplacian filter is entrenched. Laplacian is a derivative operator which sharpened the image but drives constant areas to zero. Hereby adding the original image back restored the gray level tonality of the input image. Hence the image is prepared for segmentation of WBC from the image background.

The intensity inconsistency in each region of a cell is the biggest problem in the cell segmentation and classification, particularly in gray-scale images. In [3], Nipon applied a 15x15 median filter is used to ease the problem.

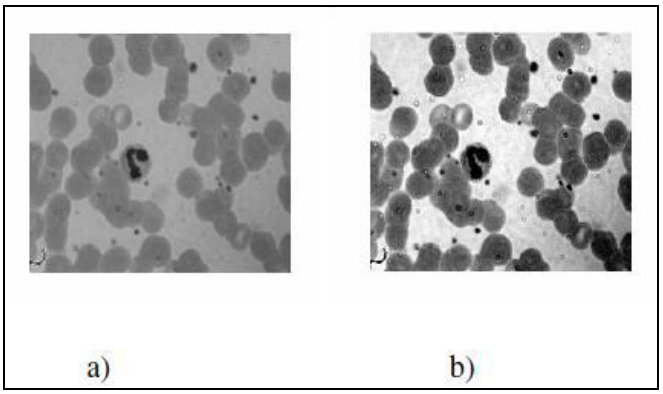

Fig.2: a) A typical image as taken b) after contrast enhancement

2.2. Segmentation

To segment the desired WBC object from the background it is found that the red component of the RGB input image gives the best contrast between the background and the blood cells components including WBC, RBC and platelets as shown in Fig. 3. However, when the blue channel is used, WBC fades out while with the green channel the WBC cytoplasm color is close to the RBCs color. In order to produce a representative binary image, Otsu's adaptive thresholding algorithm (Otsu, 1979) is then applied on the red channel. Watershed distance transform [11] process is applied on the binary image obtained by Otsu's thresholding [11].

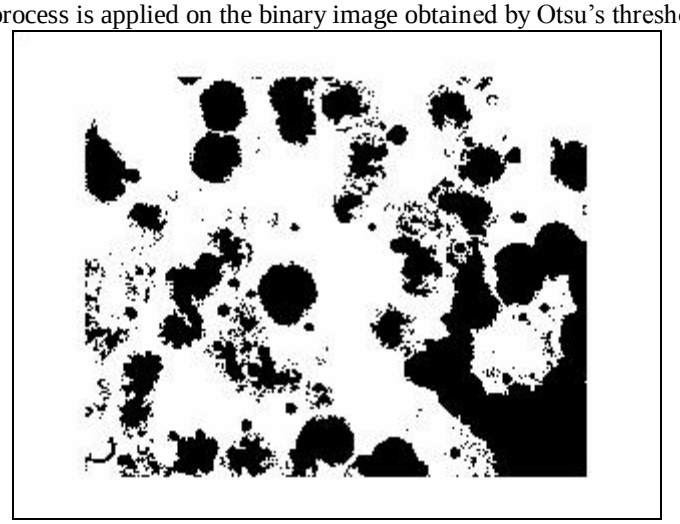

Figure 3: Red channel image after applying Otsu's adaptive thresholding algorithm

Using watershed on blood slides, different objects (including WBC, RBC, platelets and stain), are extracted from the original image. Fig.4 shows the results of the watershed algorithm when applied on the image. As it can be seen, a considerably large number of segments are achieved while the image has only one WBC. At this stage a decision has to be made in order to figure out which masks, of the many obtained from watershed, represent WBC. Each area obtained with the watershed was singled out by masking it with the original image. Furthermore, a bounding box, to reduce the background part and better prepare the segment object for further processing, is created. The first stage of WBC classification is based on the area of the bounding box [11] representing each element obtained from watershed stage. Since WBC has distinctive area size when compared with other elements a threshold value is obtained, using trial and error, below which a segment is rolled out of being a WBC.

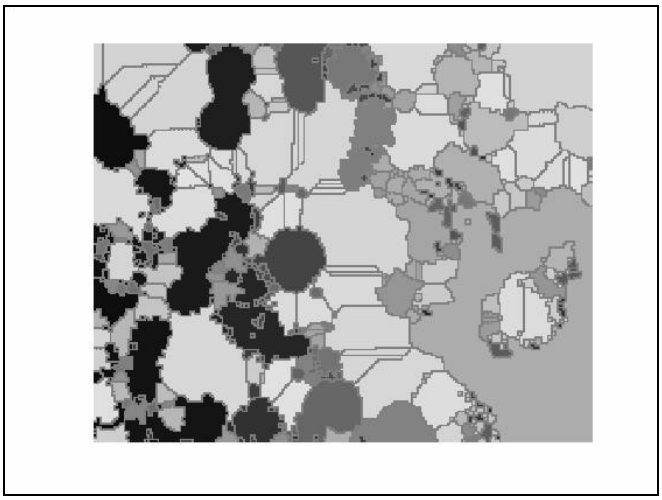

Fig.4: Watershed output of the image

The other two features that are used to classify WBC from other objects were color averaging (A) and the aspects ratio (AR) of the bounding box given as:

$$A = \frac{\Sigma_{x=0}^{N}\Sigma_{y=0}^{M}R(x,y) + \Sigma_{x=0}^{N}\Sigma_{y=0}^{N}G(x,y) + \Sigma_{x=0}^{N}\Sigma_{y=0}^{M}B(x,y)}{3} \quad (2.1)$$

$$AR = \frac{N}{M} \quad (2.2)$$

N and M are the width and height of the bounding box in pixels of the region under consideration. R, G and B refer to RGB individual matrix values. Threshold values for the above two features are obtained for typical WBC, and any bounding box that does not meet these thresholds conditions is then disregarded as being a WBC.In [8], the marker controlled watershed defines an internal and external seeds for the separation of WBC and background. From the segmented WBC the nucleus and cytoplasm is again separated using histogram thresholding approach. The fuzzy c-means (FCM) algorithm [3], as a typical clustering algorithm, has been utilized in a wide range of engineering and scientific disciplines such as medical imaging, bioinformatics, pattern recognition, and data mining. In [5] segmentation using morphological operations of watershed transformation is proposed.

KFCM [1] adopt a new kernel induced metric in the data space to restore the original Euclidean norm metric in FCM. Due to the fuzzy measures involved, the algorithm takes much computation time. The method was used to segment acute lymphocytic leukemia images and performed normal lymphocyte and lymphoblast classification.

2.3. Feature extraction

The most important problem in generation of features of blood cells is to characterize them in a way enabling the recognition of different blast types with the highest accuracy. The features to be extracted from nucleus are geometrical features like area, radius, perimeter, symmetry, concavity, compactness, solidity, eccentricity, elongation, form factor, Texture Features which includes homogeneity, energy, correlation, entropy, contrast, angular second momentum, Color Features like mean color values, Statistical Features like mean value, variance, skew-ness, kurtosis of the histograms of the image matrix and the gradient matrix for RGB or HSV or L*a*b color space (whichever appropriate) [11].

In [8], eight shape features are extracted from the nucleus of WBC and a ratio between the area of nucleus and cytoplasm is also taken into account. Shape features are area, perimeter, compactness, eccentricity, orientation, solidity, form factor and roundness. General shape features are extracted using the standard procedure present in the MATLAB Image Processing Toolbox (7.6). Form-factor and Roundness are calculated as follows,

$$Formfactor = \frac{4\pi Area}{Perimeter^2} \quad (2.4)$$

$$Roundness = \frac{4Area}{\pi(majoraxislength)^2} \quad (2.5)$$

To demonstrate the initial application of the automatic nucleus-segmented images, [3] calculated the area of nucleus in each image to be the feature to the Bayes classifier. Fabio [4] extracted features regarding the gray-level intensity pattern of the image (i.e., granularity of the color, uniformity), morphological features such as the perimeter, the area, the momentums of the image, etc.

Recognition of the blood cell on the basis of its image needs generation of the numerical features well describing the differences of images belonging to different classes. In characterizing the images by the numerical values, get the features strictly corresponding to these on the basis of which the human expert makes his diagnosis, that is the geometry of cell, texture, color and intensity of the image associated with different cell types. Four families of features have been created in this way [5]. The geometrical features include such parameters as radius, perimeter, area, the area of convex part of the cell, compactness, concavity, symmetry, major and minor axis lengths, etc.

These parameters are determined only for the nucleus of the cell. The texture [5] refers to an arrangement of the basic constituents of the material and in the digital image is depicted by the interrelationships between spatial arrangements of the image pixels. Up to 105 texture features have been generated for the cell image at normal and reduced resolutions.

The next set of features [5] has been generated from the analysis of the intensity distribution of the image. The histograms of the image and gradient matrix of such intensity have been determined for R, G, B components of the image. On the basis of such analysis, generated features like mean and variance of the histogram of the image of nucleus and cytoplasm (separately) as well as for the gradient matrix of the image, the skew-ness and kurtosis of the image of the whole cell as well as for the gradient matrix of the whole cell. Up to 24 statistical features have been generated in this way for two colors (red and green).

The last set of features [5] is related to the morphological operations performed on the image (erosion, dilation, opening and closing). These parameters include the area and number of separated objects of the image after application of some morphological operations. Up to 16 morphological parameters have been generated in this way. All features have been normalized, dividing their original values by the corresponding maxima.

2.4. Classification

Classification is the task of assigning to the unknown test vector to a known class. In [11], a reinforcement learning algorithm is proposed. The RL approach will classify the types of leukemia into ALL, AML, CLL and CML.

For classification Naive Bayes classifier was chosen for differentiation of five types of WBC nuclei depending on those four shape features [8]. Naïve Bayes classifier is a simple probabilistic classifier which is based on Bayesian statistics with strong independence assumptions. Simply, a Naive Bayes classifier express that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature and classifier efficiency depends upon all properties of that particular feature which are independently contribute to the probability. This classifier is trained with supervised learning and its parameter estimation can be based on maximum likelihood scheme.

In [5], the numerical experiments of recognition of 10 classes of blood cells typical for leukemia have been performed using SVM of Gaussian kernel and one against one mode of operation. The problem is really difficult, because of large number of recognized classes and also of close similarity of the representatives of the cells belonging to different classes. At the same time there is a large variation among cells belonging to the same family of cells.

III. Conclusion

Different techniques used in various phases of Leukemia detection were discussed. From the study, those methods felt to be better can be combined to produce a new system for leukemia detection. It is proposed that the Marker controlled watershed algorithm [5] yield better result in segmentation part. Fuzzy c-means clustering [1] also produces satisfactory result. But computation time is high. As more features are extracted, better will be the result. For that, Gray level co-occurrence matrix can be used.

REFERENCES

[1]. Rubhala, "Image segmentation of leukemia using kernel based fuzzy c means clustering", International Journal of Communications and Engineering Volume 03 No.3, Issue: 03 March2012
[2]. Juma Al-Muhairy, Yousef Al-Assaf, "Automatic white blood cell segmentation based on image processing", 2005 IFAC
[3]. Nipon Theera-Umpon, "Patch-Based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering", ECTI transactions on electrical eng., electronics, and communications vol.3, no.1 February 2005
[4]. Fabio Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images", CIMSA 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Giardini Naxos, Italy, 20-22 July 2005
[5]. T. Markiewicz and S. Osowski, "Data mining techniques for feature selection in blood cell recognition", ESANN'2006 proceedings of Symposium on Artificial Neural Networks Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.
[6]. Nipon Theera-Umpon, "Morphological Granulometric Features of Nucleus in Automatic Bone Marrow White Blood Cell Classification", IEEE transactions on information technology in biomedicine, VOL. 11, NO.3, MAY 2007

[7]. S. H. Rezato_ghi H. Soltanian-Zadeh R. Shari_an R.A. Zoroo_, "A New Approach to White Blood Cell Nucleus Segmentation Based on Gram Schmidt Orthogonalization", International Conference on Digital Image Processing, 2009 IEEE
[8]. Madhumala Ghosh', Devkumar Das, Subhodip Mandal, Chandan Chakraborty, Mallika Pal, Ashok K Maity, Surjya K PaIn, "Statistical Pattern Analysis of White Blood Cell Nuclei Morphometry", Proceedings of the 2010 IEEE Students' Technology Symposium 3-4 April 2010, IIT Kharagpur.
[9]. P.S.Hiremath, "Automated Identification and Classification of White Blood Cells (Leukocytes) in Digital Microscopic Images", IJCA Special Issue on Recent Trends in Image Processing and Pattern Recognition, RTIPPR, 2010.
[10]. Cancer facts and figures 2013, American cancer society.
[11]. Fauziah Kasmin, Anton Satria Prabuwono, Azizi Abdullah, "Detection of leukemia in human blood sample based on microscopic images: a study", Journal of Theoretical and Applied Information Technology, 31st December 2012. Vol. 46 No.2.