

# AUTOMATED DETECTION OF WHITE BLOOD CELL CANCER DISEASES USING CLASSIFICATION TECHNIQUES – A SURVEY

Manisha. S<sup>1</sup>, Dr. T. Sethukarasi<sup>2</sup>

<sup>1</sup>Master of Engineering (ME), Department of Computer Science and Engineering, RMK Engineering College, Chennai, India

<sup>2</sup>Professor, Department of Computer Science and Engineering, RMK Engineering College, Chennai, India

\*\*\*

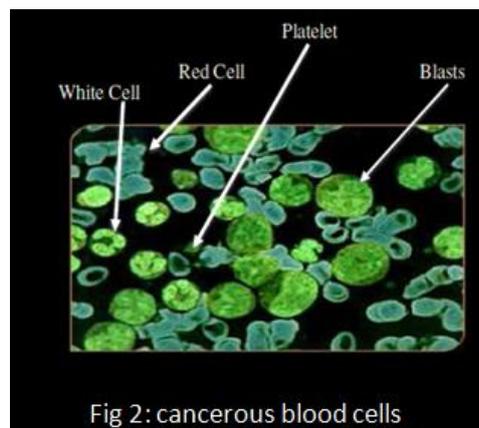
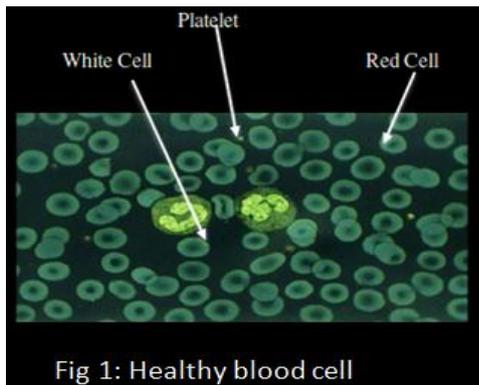
**Abstract** - Computerized analysis of white platelets tumor maladies such as Leukemia and Myeloma is a testing biomedical investigate point. Early detection plays a major role for averting complications in patient's health. Blood smear pictures are proficient and genuine wellspring of data in extraction of features in case of an occurrence of cancer growth discovery. In appraisal of cancer growth discovery like leukemia cancer image, feature extraction and reduction must be done effectively. Blood cancer can be diagnosed manually by Bone marrow biopsy, Lymph node biopsy and Lumbar puncture. Manual detection of white blood cancer cells can lead to misdiagnosis and high computation time that may even cost the patient's life. Hence, automatic detection is used for detection in less time and reduced misdiagnosis rates. In this paper, we have briefly reviewed, surveyed and analyzed the existing algorithms and methodologies used for detecting blood cancer.

**Key Words:** white blood cell, platelets, red blood cells, leukemia, myeloma, lymphoma, machine learning.

## 1. INTRODUCTION

The main purpose is to develop an automated classification system, which segments and classifies microscopic blood images from cancer and non-cancer people. The three types of blood cancer are leukemia, lymphoma and myeloma. Computerized analysis of white platelets tumor ailment such as, Leukemia and Myeloma is a testing biomedical look into point. According to the World Health Organization (WHO) [1], Cancer is considering as the second leading cause of death in the world and has claimed the lives of 8.8 million people in 2015, attributed to the death of nearly one in six deaths worldwide. Discovering those diseases in the early stages highly affects the treatment period. In addition, some of the diseases' sub-types are really confusing to the Doctors. Nowadays, there is a great tendency for diagnostic pathology to heavily rely on automated systems which can aid in the diagnosis [2]. White blood cells cancer diseases; Leukemia and Myeloma, threaten people's life nowadays. Leukemia is found when the bone marrow produces abnormal white blood cells, which don't function properly. It may be

either acute or chronic. There are four main types of leukaemia: Acute myeloid leukaemia (AML), Acute lymphoblastic leukaemia (ALL), Chronic myeloid leukaemia (CML) and Chronic lymphocytic leukaemia (CLL). Acute Myeloid Leukemia (AML) is sub categorized to (M0, M1, M2, M3, M4, M5, M6, M7). Acute lymphoblastic leukemia (ALL) is sub categorized to (L1, L2, L3). The types of acute myelogenous leukemia included: Myeloblastic (M0) on special analysis, Myeloblastic (M1) without maturation, Myeloblastic (M2) with maturation, Promyelocytic (M3), Myelomonocytic (M4), Monocytic (M5), Erythroleukemia (M6), Megakaryocytic (M7) [23]. The Symptoms of ailment ALL and AML usually like different sicknesses indications, so the analysis is very troublesome. One phase of leukemia determination systems concerns examinations in the blood utilizing a magnifying instrument. Myeloma influences plasma cells, a kind of develop white platelet in blood that produces antibodies to enable battle to off diseases. In myeloma, B lymphocytes don't develop legitimately into plasma cells yet keep on repeating quickly. These youthful cells aggregate in bone marrow and keep it from delivering different sorts of platelet. It additionally expands synthetic compounds that breakdown and redesign bone, prompting torment and cracks at times. Myeloma is frequently alluded to as different myeloma since this diminishing of bones can happen anyplace in body where there is bone marrow. Myelomas can be gathered into three phases dependent on the level of two proteins (called beta2 microglobulin and albumin) in blood. Figure 1 depicts a healthy blood cell with good quantity of platelet, red and white cell. A cancerous cell with enormous amount of white blood cells and blasts is shown in figure 2.



In this paper, we compare a portion of the strategies for automatic detection of white blood cancer cells on the basis of parameters like: sensitivity, precision, severity, efficiency, type and accuracy. The objective of this paper is to survey the important writing in the field of automated location of white blood tumor cells and to furnish looks into with a point by point asset of the accessible methodologies utilized. The paper is structured as follows. Section 2 describes the general workflow used in detection of automated white blood cancer cells. The existing methodologies in the form of literature review is given in section 3. The accuracy techniques are elaborated in section 4, followed by conclusion in section 5.

## 2. GENERAL WORKFLOW FOR AUTOMATED WHITE BLOOD CANCER CELL DETECTION

Image pre-processing, feature extraction and classification plays major role in detection of white blood cell cancers. The removal of noise is another main factor that yields accuracy to final result. The following flowchart depicts the basic steps involved in automated detection of white blood cancer cells using machine learning and image processing techniques.

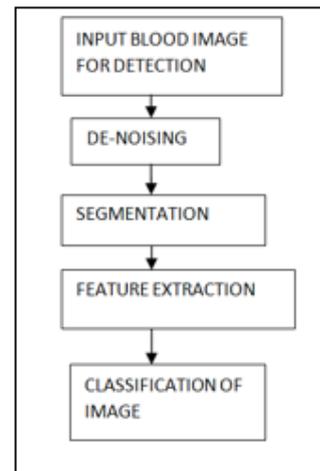


Fig -1: Block diagram of the proposed system

### 2.1 INPUT IMAGE

This stage is also known as image acquisition. For detecting, blood cancer diseases like leukemia or myeloma, input images from microscope is needed. This informational index or data-set comprises minuscule pictures of blood smear, particularly intended for the assessment and the correlation of algorithms for segmentation and image classification.

### 2.2 IMAGE PRE-PROCESSING

The main goal is to de-noise the input image. Pre-processing also includes color relevance wheel RGB images are converted to Grey color space images.

### 2.3 SEGMENTATION OF THE IMAGE

It is the process of apportioning an advanced picture into numerous portions. This is one of the most crucial steps which greatly influence on the accuracy of cancer cell detection.

### 2.4 FEATURE EXTRACTION

This method used to change the information into set of highlights and is a type of dimensionality decrease. The feature set will separate the critical data from the input information if the highlights are extricated effectively. Some of the features are Haar wavelet features, Hausdorff dimension with and without LBP, Shape features, Gray level co-occurrence features, Haralick's texture features or GLCM, Color feature.

## 2.5 CLASSIFICATION

The final stage in the process is classification. It helps to determine the existence of white blood cancer cell in blood image. It includes several algorithms like random forest, support vector machines, nearest neighbour, neural networks, decision trees.

## 3. LITERATURE REVIEW ON EXISTING METHODOLOGIES USED IN AUTOMATED CANCER CELL DETECTION TECHNIQUES

Automated detection is very useful in diagnosing the white blood cancer cells. With the help of automated system, the work of oncologists can be accurate that helps to know the current stage of patient's health and the cost for detection can also be reduced. The rate of misdiagnosis can be greatly reduced and early detection of cancer can be achieved. The existing systems on automated cancer cell detection mainly work on image pre-processing, segmentation, feature extraction and classification. Components of Medical image processing includes **Acquisition of image, improvisation, Representation, Transformation, Restoration, processing of colour images, compression, image processing, morphological image processing, segmentation and object identification/ recognition.**

This writing talks about strategies for the discovery of leukemia through image processing methods. As a rule, the Strategy of distinguishing leukemia through microscopic images of a blood test was separated into three phases:

- a. Cell Image Segmentation
- b. Feature Extraction of Cell Image
- c. Recognition and Classification of malignant growth Cell.

Hend Mohamed, Rowan Omar, Nermeen Saeed, Ali Essam, Nada Ayman, Taraggy Mohiy and Ashraf AbdelRaouf [1] proposed a hybrid methodology to detect white blood cancer cells using deep learning with support to learning system. Gaussian distribution was used for segmentation and random forest as classifier. With a count of 105 images as data set, an accuracy of 93% for L1, L2, M5 and 95% for L3, M2, M3 and Myeloma were achieved.

Anitta K Varghese et al., [4] proposed an efficient methodology with the advantage of, its simplicity, classification of complete blood smear images and helps

to segment and detect nucleated cells. Division is finished utilizing K-mean division and layer subtraction division. Hausdorff measurement were utilized for dimensionality decrease. Direct SVM two-class classifier is utilized on the grounds that it is modest, and gives a decent execution.

Iterative Thresholding calculation is utilized for division reason particularly from loud pictures by Niranjana Chatap, Sini Shibu [5] was proposed. This calculation conquers the issue of cell extraction and division from substantial uproarious pictures. Morphological way to deal with cell picture division is more exact than the established watershed-based calculation. A basic thresholding approach is connected and the calculation is determined about blood smear pictures from priori data.

W., Qiang, Zhongli, Z., [6] proposed a reinforcement learning algorithm. This approach classifies the types of leukemia into Acute myeloid leukaemia (AML), Acute lymphoblastic leukaemia (ALL), Chronic myeloid leukaemia (CML) and Chronic lymphocytic leukaemia (CLL).

S. Mohapatra et al., [7] in the year 2010 proposed Unsharp Masking Sub-Imaging Bounding Box as selective filtering and used Fuzzy C-Means Clustering followed by Nearest Neighbour Classification using support vector machine(SVM).

Taha J. Alhindi et al., [8] proposed a novel method for extracting features using local binary patterns, the histogram of gradients, and a pre-trained deep network. Three common image classification methods, including support vector machines, decision trees, and an artificial neural network was used. KIMIA Path960, a freely accessible dataset of 960 histopathology pictures extricated from 20 diverse tissue scans to test the exactness of classification and highlight or feature extractions models utilized in the examination, explicitly for the histopathology pictures. SVM achieves the highest accuracy of 90:52% using local binary patterns as features which surpasses the accuracy obtained by, namely 81:14%.

Karthikeyan and Poornima [9] proposed a methodology on automated detection of leukemia using Adaptive median filter for noise elimination and adaptive Histogram Equalization for contrast enhancement in preprocessing stage. They applied k-means and Fuzzy c-means clustering for segmentation. They employed Support Vector Machine (SVM) for classification. 90% accuracy was achieved by using k-means using fuzzy logic.

An another method [10] to detect acute leukemia was based on Otsu's segmentation method and feature extraction by contour signature proposed by Tejashri G. Patil, V. B. Raskar.

Mu-Chun Su et al., [11] 2014 proposed a work with title "A Neural- Network- Based Approach to White Blood Cell Classification" that incorporates HSI Color Space for preprocessing and discriminating region morphological operators and 7x7 median filter for image postprocessing for segmenting five types of WBC. Classification techniques includes Multilayer Perceptron(MLP), SVM and Hyperrectangular Composite Neural Networks for classifiers (HRCNN). An accuracy of 99.11% was achieved by MLP followed by 97.55% by svm and 88.95 % by HRCNN classifiers.

Khot S.T et al[12], utilized Support Vector Machine(SVM) classifier to distinguish leukemic cells. The highlights are separated from the picture and connected to classifier.

Carolina Reta, Leopoldo Altamirano, Jesus A. Gonzalez, Ranquel Diaz and Jose S. Guichard [13] presented a two phase methodology to analyze white blood cancer cell using image processing and data mining techniques. A novel method used for segmenting color and texture information was proposed using Markov random field. CIE L\*a\*b\* color space were employed for color features, instance base classifier, decision trees were used for classification.

Sarrazfzadeh et al. [14] proposed an another methodology concentrated fundamentally on separate between M2, M3 and M5 sub-sorts to assess their presented strategy. They utilized a dataset made out of 27 minuscule pictures of three sub-types of AML; 9 AML of M2, 10 AML of M3 and 8 AML of M5. The methodology was connected in the L\*a\*b\* shading space. Division is performed utilizing K-means clustering to isolate leukocytes from other blood segments. Features are extricated with the end goal to be classified using Discriminative Dictionary Learning (DDL). An accuracy of 97.53% was achieved on Medical Image and Signal Handling Research Centre (MISP) dataset.

[15] K-Means, Fuzzy C-Means, Moving K-Means, seeded region growing area was proposed by Harun et al. 2015.it has also achieved accuracy of 98.2% when SVM was used as image classifier.

Piuri et al.[16] , paper centres around the issue of Identification and arrangement of white platelets by magnifying instrument pictures. This paper shows a procedure to accomplish a completely mechanized discovery and order of leucocytes by magnifying lens shading pictures recognizing the accompanying classes:

Basophil, Eosinophil, Lymphocyte, Monocyte and Neutrophil. The proposed framework right off the bat individuates in the blood picture the leucocytes from the others platelets, also it removes morphological files lastly it groups the leucocytes by a neural classifier.

Sinha and Ramakrishnan (2002) [17] proposes A two-part approach for identifying cytoplasm and nucleus of WBC in colour images using HSV colour space, k-means, EM algorithm that achieved an accuracy of 80%.

Himali P. Vaghela et al[18], introduced shape based features finding is more exact than watershed transform, K means clustering, and histogram equalizing methods for counting leukemic cells checking leukemic cells and it likewise gives most astounding exactness 97.80%

A system proposed by Bhattacharjee and Saini [19] for ALL detection using Watershed Transformation Technique. They connected difference improvement and quality modification for upgrading pictures prior to division. In division, they utilized watershed calculation, disengaging the platelet and the cell core. Classification was done using Binary Search Tree (BST) achieving 86% and Gaussian Mixture Models (GMM) achieving 93% .

A technique by Agaian et al. [20] was proposed that automatically detects and segments AML in blood smears. Segmentation was done in the CIELAB Color space by K-Means clustering algorithm. Hausdorff Dimension features were computed using the box counting method and Local Binary Pattern (LBP). Using Support Vector Machine (SVM), Classification was achieved 98% accuracy.

#### 4. PERFORMANCE MEASURES

In this section, the evaluation approaches and the performing measures are described. S. AGAIAN ET AL. (2014), evaluated the performance with the end goal to guarantee the adequacy of our framework, utilized certain measures (valuation parameters) in order to make the decisions. These measures are: precision, specificity, sensitivity, F-measure and accuracy. When attempting to classify a specimen, there are four outcomes: **True Positives (TP)**, when cancer cells are accurately distinguished; **False Positives (FP)**, when non-malignancy cells are recognized as dangerous; **True Negatives (TN)**, when non-carcinogenic cells are effectively distinguished; and **False Negatives (FN)**, when malignancy cells are recognized as noncancerous. These parameters are Computed, the general exactness of tried Image which is detailed as specified as appeared in Table I that higher estimations of these parameters will prompt better precision.

**Table -1:** Performance evaluation parameters

PARAMETERS	DEFINITION	FORMULAE
<b>Sensitivity (recall)</b>	Test's ability to identify positive results.	$Se=(TP/(TP+FN))$
<b>Precision</b>	proportion of cases with positive Results, which are correctly identified.	$P=(TP/(TP+FP))$
<b>Specificity</b>	Test's ability to identify negative results.	$Sp=(TN/(TN+FP))$
<b>F-measure</b> (overall classification performance)	It is the harmonic mean of precision and Sensitivity.	$F=(2*Sp*P)/(P+Se)$
<b>Accuracy</b>	Measure of the degree of closeness of a measured or calculated value to its actual value.	$A=((Tp+Tn)/(Tp+Tn+Fp+Fn))$

## 5. CONCLUSION

This paper briefly analyses and reviews the related work carried out on the grounds of automated blood cancer cell detection using machine learning and image processing techniques. The utilization of image processing can accelerate the procedure of leukemia detection. Image processing strategy can defeat the subjectivity of the administrator's magnifying lens, for the most part when the investigation was impacted by the level of involvement and wellness level of the spectator. The automatic detection on white blood cancer cells possesses various challenges. The automated system should correctly identify the intended objective at the earliest. Most of the methodologies uses Support vector machine (SVM) as classifiers and k-means as segmentation algorithms. The accuracy of the automated system greatly depends on type of segmentation algorithm employed, the feature to be extracted and the classification methodology incorporated.

## REFERENCES

- 1) Hend Mohamed, Rowan Omar, Nermeen Saeed, Ali Essam, Nada Ayman, Taraggy Mohiy and Ashraf AbdelRaouf "Automated Detection of White Blood Cells Cancer Diseases", in IEEE 2018 First International Workshop on Deep and Representation Learning (IEDRL).
- 2) J. B. Henry, J. P. AuBuchon, Clinical diagnosis and management by laboratory methods, Archives of Pathology and Laboratory Medicine 121 (9) (1997) 1016.

- 3) K. M. Sensing, Cancer. URL <http://www.who.int/mediacentre/factsheets/fs297/en.html>
- 4) Anitta K Varghese and Nisha J S,"Automated Screening System for Acute Myelogenous Leukemia Detection using Layer Subtraction",in:International Journal of Current Engineering and Technology ,Available online 02 Oct 2015, Vol.5, No.5 (Oct 2015)
- 5) Niranjana Chatap, Sini Shibu " Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour" (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep - Oct. 2014), PP 79-87.
- 6) W., Qiang, Zhongli, Z., 2011. "Reinforcement Learning,Algorithms and Its Application",International Conference on Mechatronic Science, Electric Engineering and Computer,,Jilin, China, August 19-22, 2011, pp. 143-1146.
- 7) S. Mohapatra, S. S. Samanta, D. Patra, S. Satpathi, Fuzzy based blood image segmentation for automated leukemia detection, in: Devices and Communications (ICDeCom), 2011 International Conference on, IEEE, 2011, pp. 1-5.
- 8) Taha J. Alhindi, Shivam Kalra, Ka Hin Ng, Anika Afrin, Hamid R. Tizhoosh ,"comparing LBP,HOG and deep features for classification of histopathology images", in: IEEE world congress on computational intelligence (IEEE WCCI),2018.
- 9) T. Karthikeyan, N. Poornima, Microscopic image segmentation using fuzzy c means for leukemia diagnosis, Leukemia 4 (1).
- 10) Tejashri G. Patil, V. B. Raskar,," Blood microscopic image segmentation & acute leukemia detection"(IJERMT) ISSN: 2278-9359 (Volume-4, Issue-9)SEP 2015.
- 11) M. Su, C. Cheng, P. Wang. "A Neural-Network-Based Approach to White Blood Cell Classification." The Scientific World Journal Volume 2014, Article ID 796371. 2014.
- 12) Khot S.T, Sneha Bhalekar, Divya Jaggi and Dolly Rani, "An innovative approach in myelogenous leukemia detection using attribute analysis" (IJAE) ISSN (Print): 2278-8948, Volume-2, Issue-5, 2013.

- 13) Carolina Reta, Leopoldo Altamirano, Jesus A. Gonzalez, Ranquel Diaz and Jose S. Guichard (2015), "segmentation of bone marrow cells images for morphological classification of acute leukemias". Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010).
- 14) O. Sarrafzadeh, H. Rabbani, A. M. Dehnavi, A. Talebi, Detecting different sub-types of acute myelogenous leukemia using dictionary learning and sparse representation, in: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, 2015, pp. 3339-3343.
- 15) N.H. Harun, Nasir, A.S. Abdul, Mashor, M.Y. Mashor, R. Hassan, "Unsupervised Segmentation Technique for Acute Leukemia Cells Using Clustering Algorithms." International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9, No:1, 2015.
- 16) V. Piuri, F. Scotti: "Morphological Classification of Blood Leucocytes by Microscope Images" Proc. International Symposium on Computational Intelligence for Measurement Systems and Applications, Boston, MD, USA, July 2004.
- 17) N.Sinha, A.G.Ramakrishnan,"automation of differential blood count",2003
- 18) Himali P. Vaghela, Hardik Modi, Manoj Pandya, Manoj Pandya," Leukemia Detection using Digital Image Processing Techniques" (IJ AIS) - ISSN : 2249-0868 Volume 10 - No.1, November 2015
- 19) R. Bhattacharjee, L. M. Saini, Detection of acute lymphoblastic leukemia using watershed transformation technique, in: Signal Processing, Computing and Control (ISPC), 2015 International Conference on, IEEE, 2015, pp. 383-386.
- 20) S. Aghaian, M. Madhukar, A. T. Chronopoulos, Automated screening system for acute myelogenous leukemia detection in blood microscopic images, IEEE Systems journal 8 (3) (2014) 995-1004.
- 21) D.Goutam and S.Sailaja ," Classification of Acute Myelogenous Leukemia in Blood Microscopic Images using Supervised Classifier", in: 2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015, Coimbatore, TN, India.
- 22) <http://www.cancercenter.com/leukemia/types/>
- 23) [ALL-IDB] Acute Lymphoblastic Leukemia Image Database for Image Processing. Department of Information Technology - Universita degli Studi di Milano. [cited 2016 Nov
- 24) <https://www.lls.org/leukemia>
- 25) Medical image and signal processing research center. URL <http://misp.mui.ac.ir/en>
- 26) I. Steinwart, A. Christmann, Support Vector Machines, Information Science and Statistics, Springer New York, 2008.