International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

BLOOD SMEAR ANALYSIS FOR ACUTE LEUKEMIA DETECTION-MACHINE LEARNING APPROACH

BIJI G
Electrical Department,
Govt Engineering College, Barton Hill, Thiruvananthapuram,

Abstract— Whenever the normal functioning of the body or any of its part becomes impaired, diseases occur and may require medical treatment. In general, diseases can be classified on the basis of their cause and cell of origin i.e. infectious, immunological, endocrine, genetic, neoplastic, and traumatic etc. Among all diseases the quest for understanding leukemia, a malignant neoplastic disorder is in the research forefront for several investigators including biologists, clinicians, and chemists. Physicians across the globe are interested in understanding the biology of diseases, and how it can be prevented, or treated. Studies reveal that excessive workload, shortage of trained pathologist, and use of conventional hematological evaluation methods are some of the leading causes behind delayed or wrong diagnosis in India. Such shortcomings can be overcome by the utilization of quantitative microscopic techniques in the precise characterization of blood test samples facilitating early diagnosis of blood cancers. The development of an automated system for cancer diagnosis in the scanned microscopic images involves four main computational steps i.e. preprocessing, segmentation, feature extraction and detection. The final machine learning approach that locates circles with accuracy even considering complicated conditions and noisy images.

Keywords— Neoplastic, preprocessing, segmentation, feature extraction, Machine learning, Leukemia

I. INTRODUCTION

Leukemia also known as liquid cancer which develops from cells in the blood, bone marrow, and lymphatic system. It is different from other cancers as it does not produce solid masses or tumors. In leukemia, the abnormal white blood cells flood the marrow, providing no room for red blood cells and platelets. This can affect a patient in several ways i.e. decrease in red blood cells can result with anemia, drop in platelet count decreases the clotting ability of the blood. Moreover due to abnormal nature of white blood cells, they lack the ability to fight infections. The usual symptoms of leukemia include fatigue, frequent infections, and easy bruising and bleeding. Depending on the clinical course, leukemia disease can be preliminary classified as either acute with rapidly progressing disease with a predominance of highly immature blast cells, or chronic which denotes slowly progressing disease with increased numbers of more mature cells [1]. However, additional classification of leukemia are developed to further identify differences in the response to treatment, prognosis and are based on the hematopoietic cell of origin i.e. myelocytic (myeloid) or lymphocytic (lymphoid). As per World Health Organization (WHO) acute leukemia in general can be defined as malignant neoplasms with more than 20% blasts (myeloid or lymphoid) in the peripheral blood/bone marrow. In this study, we investigate on one such acute condition of malignant proliferation of

lymphoid cells known as acute lymphoblastic leukemia. Acute lymphoblastic leukemia (ALL) is a malignant disease caused by the genetic alterations of the lymphocyte precursor cells of the bone marrow. In the language of hematology precursors are also known as blasts, therefore ALL is known as acute lymphoblastic leukemia. ALL is characterized by excessive production of immature lymphocytes (lymphoblast) in the bone marrow preventing normal hematopoiesis. If untreated ALL can cause death due to crowding out normal cells in the bone marrow and by metastasizing to other essential organs through the peripheral blood. Clinically and biologically features of ALL are sufficiently distinct from its myeloid counterpart and warrant separate diagnostic and treatment protocols. Moreover, due to advances in molecular biology and treatment modalities subtype classification of ALL has become essential for prognostic assessment and suitable chemotherapy planning.

Microscopy based cytometry allows inspection of histological characteristics of lymphocyte for the diagnosis and classification of ALL. Although it is an invasive procedure, this modality provides evidence and display visual images of morphological components of cells and tissues under study. Visualization of underlying cellular components even exposes the texture content of cytoplasmic and nucleus regions of the lymphocytes. Provision to interpret morphological and textural features of cells assists in the diagnosis process, and is the motivation for visual microscopy. Hematopathologists have been using light microscopy for the visualization of cell and tissue samples from a long time. They rely on their clinical expertise while making decisions about the healthiness of the examined PBS or bone marrow biopsy samples. This includes distinguishing normal mature lymphocytes from leukemic blasts (lymphoblast) and identifying subtypes of lymphoblast using FAB classification. Nevertheless, variability in reported manual diagnosis may still occur [2,3] in all types of cancers including ALL. This could be due to, but not limited to morphological heterogeneity; noise arising due to improper staining process; intraobserver variability, i.e. hematopathologists inability to produce same reading while observing the same samples more than once and interobserver variability, i.e. difference in reading among hematopathologists. Therefore, over the few decades quantitative techniques have been developed and have taken over conventional pathological examinations in the process of cancer diagnosis [4]. Such techniques developed for computer aided diagnosis avoid unnecessary repeated biopsies, and offer a rigorous and reproducible method of clinical investigation. Currently, the challenge still remains in developing a value added diagnostic technique for early detection of diseases and reducing diagnostic error in comparison to the conventional procedures. Other than the development of automated differential counter, very limited research has been undertaken in the area of quantitative hematology. Researchers are yet to develop an integrated image processing based approach to differentiate mature lymphocytes from leukemic blasts. In addition, there is no dedicated image based method for which morphological features of lymphocytes can be used to subtype leukemic blasts based on cell lineages. Experimental studies showed that quantitative morphological features of normal and malignant blood samples have significant difference among them. Thus, such objective measurements can facilitate early and accurate diagnosis of ALL and its subtyping. In the following section, illustrate the use of image processing in hematology.

II. LITERATURE REVIEW

In last few years, various researchers have been attracted to digital pathology, and have contributed to the area of modern quantitative microscopy [6]. In the literature, most of the work done are devoted to overcome the problem of subjectivity in the visual assessment of morphological characteristics in stained cell/tissue samples. Although extensive research has been carried out to implement quantitative microscopy on histopathological images, studies on the automatic evaluation of hematological images for disease recognition and classification is limited. From the available literature on hematological image processing it is observed that most of the research done till date can primarily be categorized into three groups namely —

A1. Leukocyte or White Blood Cell (WBC) image segmentation

B1. Differential blood count
C1. Automated leukemia detection

A1. Leukocyte Image Segmentation

Leukocyte or WBC image segmentation methods available in the literature are mostly shape, threshold, region growing, or edge based schemes. Liao and Deng [7] introduced a novel WBC image segmentation scheme which is based on simple thresholding followed by contour identification. This algorithm works with an assumption that the cells are circular in shape, hence is not at all suitable for irregularly shaped lymphoblasts (malignant lymphocytes). Angulo *et al.* [8] proposed a two stage blood image segmentation algorithm based on automatic thresholding and binary filtering. This scheme exhibits good segmentation performance in terms of cytoplasm, nucleus and nucleolus extraction in lymphocyte images. All these come at the cost of higher computational time due to the two stage segmentation process. Moreover, determination of optimum threshold for initial segmentation is always difficult due to variable staining and lighting conditions. Sinha *et al.* [9] proposed an automated leukocyte segmentation scheme using Gaussian mixture modeling and EM algorithm. This method is fully unsupervised and even no parameter tuning is necessary, however this scheme does not perform well for all stains. Umpon [10] introduced patch based WBC nucleus segmentation using fuzzy clustering. Even if the nucleus segmentation is accurate, there is no provision for cytoplasm extraction which is equally important for leukemia detection. Dorini *et al.* [11] used watershed transform based on image forest transform to extract the nucleus. Concurrently, size distribution information is used to extract the cytoplasm from the background including RBC. While effective for nucleus segmentation this method fails when the cytoplasm is not round. Dorin Comaniciu *et al.* [12] proposed an efficient cell segmentation algorithm that detects clusters in the *L*u*v* color space and delineates their borders by employing the gradient ascent mean shift algorithm. Though this method is effective in accurate nucleus segmentation, there is no provision for cytoplasm extraction which is also essential for ALL detection. Yang *et al.* [13] used color gradient vector flow (GVF) active contour model for leukocyte segmentation. The algorithm has been developed in the *L u v* color space. They have incorporated color gradient and *L2E* robust estimation technique into the traditional GVF snake model. Though the segmentation performance showed promising results in comparison to the mean shift approach and the standard color GVF snake, the test data is unable to distinguish weak edges and textures, thereby limiting its ability to segment lymphocytes. Yi *et al.* proposed a PSO trained on–line neural network for WBC image segmentation. It uses mean–shift and uniform sampling for reducing the raining data set. Despite the reduction in training time, this scheme is found to be unsuitable for differentiating nucleus from cytoplasm accurately. Shitong proposed a hybrid method combining threshold segmentation followed by mathematical morphology and fuzzy cellular neural networks. However, despite high running speed and good leukocyte detection it is unable to separate cytoplasm and nucleus. Chinwaraphat *et al.* proposed a modified fuzzy c–means clustering technique. The modification is performed to eliminate false clustering due to uncertainty in determining the belongingness at the conjunction of cytoplasm and nucleus. The segmentation performance is only compared to traditional Fuzzy c–Means and manual cropping is necessary for the test images. Meurie *et al.* introduced an automatic segmentation scheme based on combination of pixel classification. However, despite hybridization of classifiers the average segmentation performance is not so high. Further the use of multiple classifiers increases the average running time. Ghosh *et al.* proposed a marker controlled watershed segmentation technique to extract the entire WBC from the background. Although the proposed technique usually performs well in extracting the WBC from the background, it obtains rather poor result while extracting cytoplasm and nucleus from the background. Determination of accurate threshold to separate nucleus from cytoplasm is important, and no specific methods has been presented for its estimation. Ghosh *et al.* proposed a threshold detection scheme using fuzzy divergence for leukocyte segmentation. Various fuzzy membership functions i.e. Gamma, Gaussian and Cauchy functions have been evaluated for the test images. While this method is able to segment the nucleus accurately, there is no provision for cytoplasm extraction which is also

an essential morphological component of lymphocytes for ALL detection. Ko *et al.* proposed a hybrid leukocyte segmentation scheme which employs stepwise merging rules based on mean shift clustering and boundary removal rules with a GVF snake model. Two different schemes are employed independently to extract the cytoplasm and nucleus of the leukocyte. However, the segmentation accuracy for cytoplasm needs further improvement and computation time has to be reduced.

III. PROPOSED METHOD

Since blood cells can be approximated with a quasicircular form, a circular detector algorithm may be handy. The problem of detecting circular features holds paramount importance for image analysis, in particular for medical image analysis [5]. The circle detection in digital images is commonly performed by the circular Hough transform [6]. A typical Hough-based approach employs an edge detector whose information guides the inference for circle locations and radius values. Peak detection is then performed by averaging, filtering, and histogramming the transform space. It also implies a high computational complexity yielding a low processing speed. The circle Hough Transform is a basic technique used in Digital Image Processing, for detecting circular objects in a digital image. The circle Hough Transform is a feature extraction technique for detecting circles. It is a specialization of Hough Transform. The purpose of the technique is to find circles in imperfect image inputs. The circle candidates are produced by "voting" in the Hough parameter space and then select the local maxima in a so-called accumulator matrix. In a two-dimensional space, a circle can be described by:

$$(x-a)^2 + (x-b)^2 = r^2$$

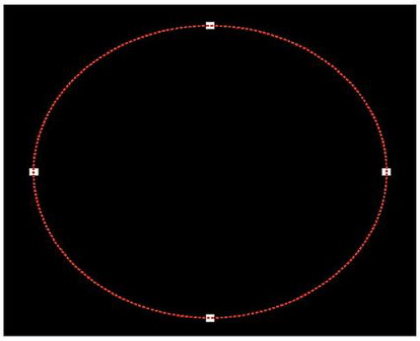
where (a,b) is the center of the circle, and r is the radius. If a 2D point (x,y) is fixed, then the parameters can be found according to (1). The parameter space would be three dimensional, (a, b, r). And all the parameters that satisfy (x, y) would lie on the surface of an inverted right-angled cone whose apex is at (x, y, 0). In the 3D space, the circle parameters can be identified by the intersection of many conic surfaces that are defined by points on the 2D circle. This process can be divided into two stages. The first stage is fixing radius then find the optimal center of circles in a 2D parameter space. The second stage is to find the optimal radius in a one dimensional parameter space. Thus the parametric space for a circle will belong to R3. As the number of parameters needed to describe the shape increases as well as the dimensions of the parameter space R increases so do the complexity of the Hough Transform. So with the Circle Hough Transform, we expect to find triplets of (x, y, R) that are highly probably circles in the image. First find out the edges using Canny edge detection algorithm. For each edge point, draw a circle assuming that edge point as center with the desired radius. The circle is drawn in the parameter space (fig 1). The 'a' value is the x-axis, 'b' value is the y-axis while the z-axis is the radii in the parameter space of the circle drawn. At the coordinates which belong to the perimeter of the drawn circle we increment the value in our accumulator matrix which essentially has the same size as the parameter space. In this way for each edge point in the input image drawing circles with the desired radii and incrementing the values in our accumulator. When completed with every edge point and every desired radius, turn attention to the accumulator. The numbers in the accumulator will represent the number of circles passing through the individual coordinates.

1) Parameters with known radius R

If the radius is fixed, then the parameter space would be reduced to 2D (the position of the circle center). For each point (x, y) on the original circle, it can define a circle centered at (x, y) with radius R according to (1). The intersection point of all such circles in the parameter space would be corresponding to the center point of the original circle.

Consider 4 points on a circle in the original image shown in fig 1.1. The circle Hough transforms is shown in fig 1.2. Note that the radius is assumed to be known. For each (x,y) of the four points (white points) in the original image, it can define a circle in the Hough parameter space centered at (x, y) with radius r. An accumulator matrix is used for tracking the intersection point. In the parameter space, the voting number of points through which the circle passing would be increased by

one. Then the local maxima point (the red point in the center in the right figure) can be found. The position (a, b) of the maxima would be the center of the original circle.



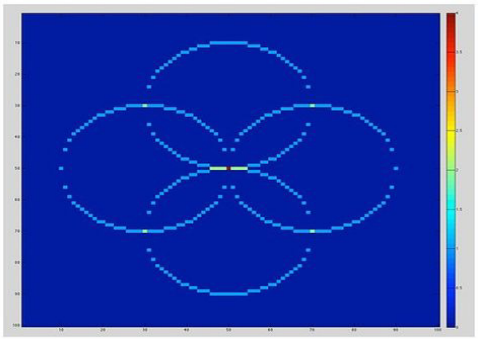
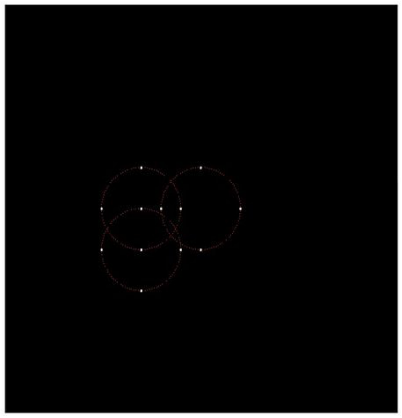


Fig 1.1.Original Image *Fig.1.2 Hough transform Image*

2)Multiple circles with known radius R

Multiple circles with same radius can be found with the technique Accumulator matrix and voting.In practice, an accumulator matrix is introduced to find the intersection point in the parameter space. First, we need to divide the parameter space into "buckets" using a grid and produce an accumulator matrix according to the grid. The element in the accumulator matrix denotes the number of "circles" in the parameter space that passing through the corresponding grid cell in the parameter space. The number is also called "voting number". Initially, every element in the matrix is zeros. Then for each "edge" point in the original space, we can formulate a circle in the parameter space and increase the voting number of the grid cell which the circle passing through. This process is called "voting". After voting, we can find local maxima in the accumulator matrix. The positions of the local maxima are corresponding to the circle centers in the original space.



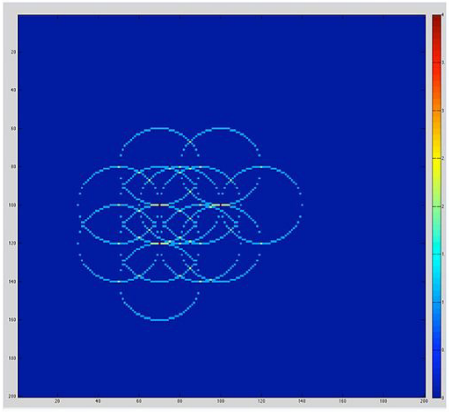


Fig 2.1.Original Image *Fig.2.2 Hough transform Image*

3) **Circle parameter with unknown radius**

Since the parameter space is 3D, the accumulator matrix would be 3D, too. We can iterate through possible radii; for each radius, we use the previous technique. Finally, find the local maxima in the 3D accumulator matrix. Accumulator array should be A[x,y,r] in the 3D space. Voting should be for each pixels, radius and θ A[x,y,r] += 1

The algorithm :

1. For each A[a,b,r] = 0;

2. Process the filtering algorithm on image Gaussian Blurring, convert the image to grayscale (grayScaling), make Canny operator, The Canny operator gives the edges on image.
3. Vote the all possible circles in accumulator.
4. The local maximum voted circles of Accumulator A gives the circle Hough space.
5. The maximum voted circle of Accumulator gives the circle.

IV. EXPERIMENTAL RESULTS

This section is to evaluate the performance of proposed method through human supervision. Fig.3(a),(b),(c) shows the real sample of microscopic images with resolutions of 1280 X 960. The resultant images after applying the initial clustering technique using Adaptive K means clustering (AKM) results are shown in 4(a), (b) and (c) respectively.

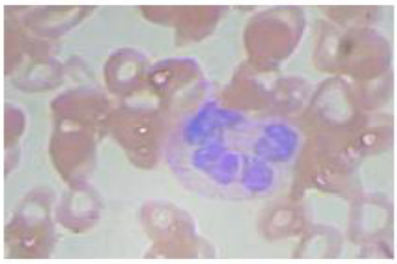


Fig 3(a) Image

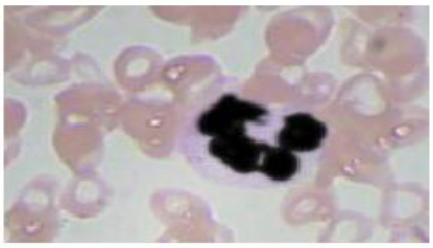


Fig 3(a.1) Segmented Image

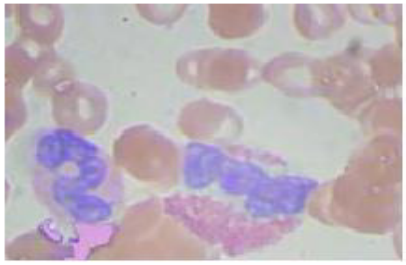


Fig 3(b) Image

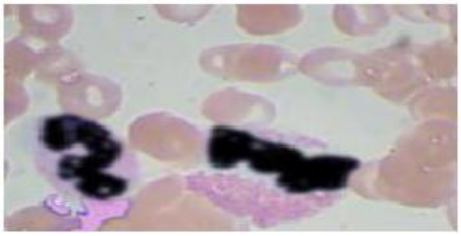


Fig 3(b.1) Segmented Image

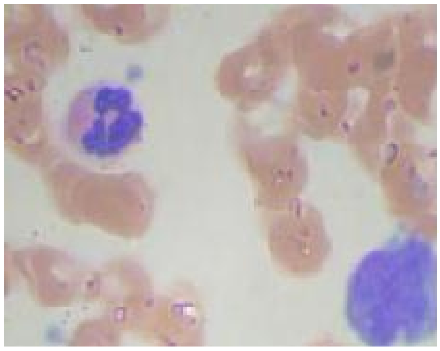


Fig 3(c) Image

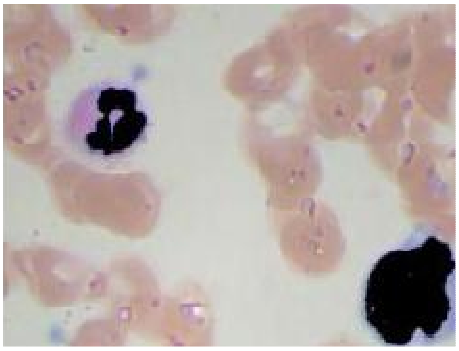


Fig 3(c.1) Segmented Image

Fig 3. Image and Segmented Image using Adaptive K Means

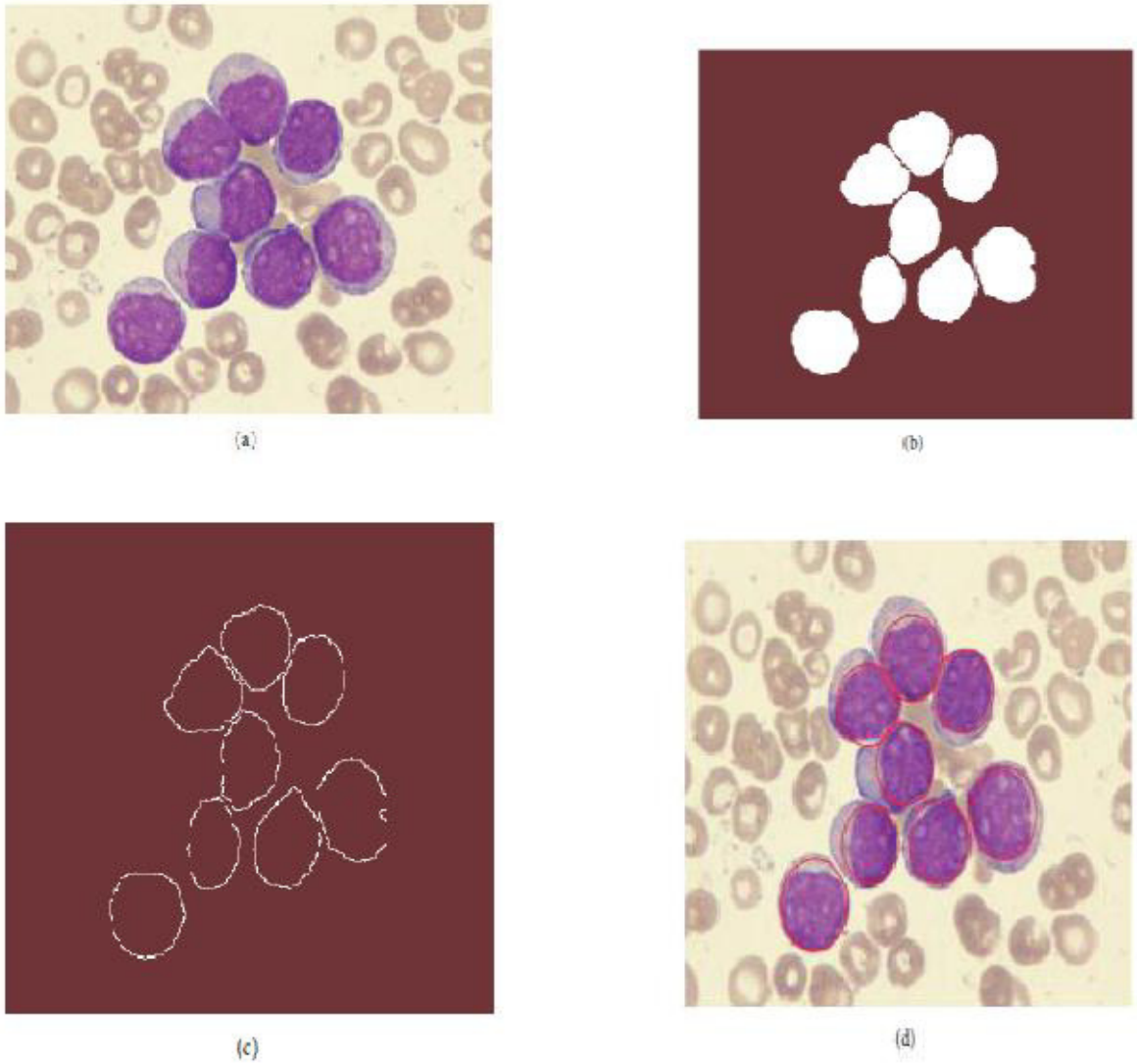


Figure 4: Resulting images of WBC detector: (a) original image, (b) image segmented by the DEM algorithm,(c) edge map, and (d) the white detected blood cells with red circles

In this experiment ,make a comparison using k-centroids value of k=6 with 10 iterations due to the convergence occurs with less than 10 For fuzziness parameter, selected the degree of fuzziness, m = 2 and threshold, = 0.01.This algorithm shows the best results.

Figure 5(a) shows an example image employed in the test. It was used as input image for the WBC detector. Figure 5(b) presents the segmented WBCs obtained by the DEM algorithm. Figures 5(c) and 5(d) present the edge map and the white blood cells after detection, respectively. The results show that the proposed algorithm can effectively detect and mark blood cells despite cell occlusion, deformation or overlapping. Other parameters may also be calculated through the algorithm: the total area covered by white blood cells and relationships between several cell sizes. Experimental tests have been developed in order to evaluate the performance of the WBC detector. It was tested over microscope images from blood smears holding a 600 × 500 pixel resolution. They correspond to supporting images on the leukaemia diagnosis. The images show several complex conditions such as deformed cells and overlapping with partial occlusions. The robustness of the algorithm has been tested under such demanding conditions.

V. CONCLUSIONS

This paper has presented an algorithm for the automatic detection of white blood cells that are embedded into complicated and cluttered smear images by considering the complete process as a circle detection problem. The approach is based on a nature-inspired technique called the Hough transform technique which is a heuristic method that follows accumulator matrix and voting. Multiple circles with same radius found with the technique Accumulator matrix and voting. In practice, an accumulator matrix is introduced to find the intersection point in the parameter space. After voting, we can find local maxima in the accumulator matrix. The positions of the local maxima are corresponding to the circle centers in the original space.Experimental results demonstrate the high performance of the proposed method in terms of detection accuracy, robustness, and stability.

V1.REFERENCES

[1] M. Wang and R. Chu, "A novel white blood cell detection method based on boundary support vectors," in Proceedings of IEEE International Conference on Systems,Man and Cybernetics (SMC '09), pp. 2595–2598, San Antonio, Tex, USA, October 2009 .

[2] J. Wu, P. Zeng, Y. Zhou, and C. Olivier, "A novel color image segmentation method and its application to white blood cell image analysis," in Proceedings of the 8th International Conference on Signal Processing (ICSP '06), November 2006.

[3] S. Wang, F. L. Korris, and D. Fu, "Applying the improved fuzzy cellular neural network IFCNN to white blood cell detection," Neurocomputing, Vol.70,No 7-9,pp 1348-1359 ,2007 .

[4] G. Karkavitsas and M. Rangoussi, "Object localization in medical images using genetic algorithms," World Academy of Science,Engineering and Technology,Vol.2,pp.499-502,2007

[5] H. Muammar and M. Nixon, "Approaches to extending the Hough transform," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89),vol. 3, pp.1556-1559,May 1989.

[6] T. Atherton and D. Kerbyson, "Using phase to represent radius in the coherent circle Hough transform," in Proceedings of IEE Colloquium on the Hough Transform, pp. 1–4,May 1993.

[7] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, no. 6, pp. 381–395, 1981.

[8] R. A. Smith, V. Cokkinides, D. Brooks, D. Saslow, and O. W. Brawley,"Cancer screening in the United States, 2010: A review of current American Cancer Society guidelines and issues in cancer screening," *CA Cancer J. Clin.*, vol. 60, no. 2, pp. 99–119, Mar./Apr. 2010.

[9] H.Xue, B. Lu, andM. Lai, "The cancer secretome:Areservoir of biomarkers," *J. Transl. Med.*, vol. 6, no. 1, pp. 6–52, Sep. 2008.

[10] H. D. Nelson, K. Tyne, A. Naik, C. Bougatsos, B. K. Chan, and L. Humphrey, "Screening for breast cancer:An update for the U.S. Preventive Services Task Force," *Annu. Int. Med*, vol. 151, no. 10, pp. 727–737, Nov. 2009.

[11] Q. B. Li,X. J. Sun,Y. Z. Xu, L. M.Yang,Y. F. Zhang, S. F.Weng, J. S. Shi, and J. G.Wu, "Diagnosis of gastric inflammation and malignancy in endoscopic biopsies based on Fourier transform infrared spectroscopy," *Clin. Chem.*, vol. 51, no. 2, pp. 346–350, Feb. 2005.
[12] A. Zwielly, S. Mordechai, I. Sinielnikov, A. Salman, E. Bogomolny, and S. Argov, "Advanced statistical techniques applied to comprehensive FTIR spectra on human colonic tissues," *Med. Phys.*, vol. 37, no. 3, pp. 1047–1055, Mar. 2010.
[13] M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, P. Gardner, and N. W. Clarke, "FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer," *Brit. J. Cancer*, vol. 99, no. 11, pp. 1859–1866, Dec. 2008.
[14] D. I. Ellis andR.Goodacre, "Metabolic fingerprinting in disease diagnosis: Biomedical applications of infrared and Raman spectroscopy," *Analyst*, vol. 131, no. 10, pp. 875–885, Aug. 2006.

