International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 01,January-2015

Classification of Acute Myelogenous Leukemia in Blood Microscopic Images using Supervised Classifier

D. Goutam
M.E Applied Electronics, Student, Department of ECE
Jaya Engineering College
Tamilnadu, India

S. Sailaja
Assistant Professor, Department of ECE
Jaya Engineering College
Tamilnadu, India

***Abstract*—Blood cancer is a form of cancer which attacks the blood, bone marrow, or lymphatic system. It is diagnosed with a blood test in which specific types of blood cells are counted by hematologist. We considered only acutemyelogenous leukemia which is one of the blood cancer type which categories under acute leukemiaand it mostly comes among adults. Need for automatic diagnosis of leukemia arises when doctors recognize cancers under a microscope which has complete manual work and it's not good for the patient. Automatic diagnosis system which helps hematologists for easier identification and early detection of leukemia from blood microscopic images which will improve the chances of survival for the patient. In this proposed system, which mainly composed of four main stages are preprocess stage, segmentation stage, feature extraction stage and classification stage respectively. This system framework consists simple and known technique such as K-mean clustering, Local Directional path (LDP), and support vector machine (SVM) respectively. The condition of a patient is shown as normal or abnormal status with the help of classifier. The overall system performance is evaluated using the defined parameters such as sensitivity, specificity, f-measure, and precision whichused for calculating the accuracy. Ninety microscopic blood images were tested, and the proposed framework managed to obtain 98% accuracy. Finally, we compare the results of some of the existing systems with our proposed system to show our achievement on accuracy.**

Keywords—Acute myelogenous leukemia , Pre-processing stage, feature extraction stage, classification stage and segmentation

I. INTRODUCTION

The term leukemia comes from the Greek word "leukos" meaning "white" and "aim" meaning "blood." It refers to the cancer of the blood or the bone marrow (where blood cells are produced). The Leukocytes which play a major role in the diagnosis of different diseases. The Extracting information from WBC is valuable for hematologists as in [1]. Diagnosing leukemia is based on the fact that white cell count are increased with immature blast cells (lymphoid or myeloid), and neutrophils and platelets are decreased as in [2], [3]. Therefore, hematologists routinely examine blood smears under microscope for proper identification and classification of blast cells. The presence of the excess number of blast cells in peripheral blood is a significant symptom of leukemia.

There are four main types of leukemia: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML), as well as a number of less common types. ALL is the most common type of leukemia in young children. CLL most often affects adults over the age of 55. AML occurs more commonly in adults than in children, and more commonly in men than women. AML is treated with chemotherapy. CML occurs mainly in adults; a very small number of children also develop this disease. Fig. 1 shows examples of leukemic cell images gathered from the web source [4].

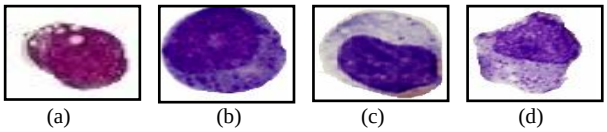
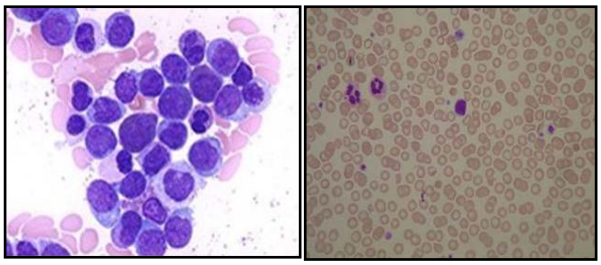


Fig. 1. Examples of leukemic blood cells showing the four different types of leukemia: (a) ALL, (b) AML, (c) CLL, and (d) CML

AML is confirmed when the marrow contains more than 30% blasts as in [2]. AML is a fast-growing cancer of the blood and bone marrow. It is fatal, if left untreated, due to its rapid spread into the bloodstream and other vital organs as in [5]. Early diagnosis of the disease is a fundamental to the recovery of patients, particularly in the case of children as in [5]. If the described symptoms are present, blood tests, such as a full blood count, renal function and electrolytes, and liver enzyme and blood count, have to be done as in [5]. Since there is no staging for AML, choosing the type of treatment can vary from chemotherapy, radiation therapy, bone marrow transplant, and biological therapy as in [6]. The Input image which is accessed from the American Society of Hematology (ASH) online image bank of leukemia cells from the web source [7]. The Fig. 2 shows two different input images, one depicting healthy cells from non-AML patients and others from AML patients.

This paper is structured as follows. Section II focuses in detail on different literature reviews. Section III shows the design methodologies being used to perform segmentation,

feature extraction, and classification. The Section IV gives detailed computer simulation process with the result for different datasets of input images. The Section V contains conclusions and future work.



(a) (b)

Fig. 2. Input Microscopy Images. (a) Myeloblasts from AML patients. (b) Healthy cells from non-AML patients.

II. PERVIOUS WORK

A number of research efforts for segmentation have been reported in the last few years. In the past, digital image processing techniques have helped to analyze the cells that lead to more accurate, standard, and remote disease diagnosis systems. However, there are a few complications in extracting the data from WBCs due to wide variation of cells in shape, size, edge, and position [1]. A detailed survey of the existing methods can be found in [2], [5, [6],[8]-[14], [16], [18], [20] and [24]-[26] and can be classified mainly into three main categories are segmentations, Features extraction and classifications.

In Medical diagnosis process one of the most important step is to segment the nucleuses from cluster of nucleus. A watershed segmentation algorithm to segment nucleus from the surrounding cytoplasm of cervical cancer images. When it comes to peripheral blood or bone marrow smears, region-based or edge-based schemes as in [8]. The merging technique for Nuclei Segmentation as in [9]. The work in [10], [11], [15]and [13] presented an unsupervised color segmentation to bring out the WBCs from acute leukemia images. The color image segmentation algorithms, it was concluded that color images present more reliable image segmentations than gray-level images which gives better segmentation but it has over-segmentation problem. The selective filtering to segment leukocytes from the other blood components and classification accuracy obtained was 95% for a different set of similar blood images using k-mean clustering as in [13]. The contour signature to identify the irregularities in the nucleus boundary with better accuracy as shown in [12].

For Feature extraction in image processing is a technique of redefining a large set of redundant data into a set of features of reduced dimension. Transforming the input data into the set of features is called feature extraction. A more robust facial and texture descriptor, named as Local Directional Pattern (LDP), was devised, where the LDP representation of face demonstrated better recognition performance than LBP as in [14]. Mostly, Local Binary Pattern (LBP) and its variants, as a feature descriptor for facial expression representationandit uses the information of intensity changes around pixels as in [19]. The Local Directional Pattern (LDP) principle which features has overcomes the limitations of LBP features as in [14], [19], and [21]. However, LDP is derived from the edge responses which are less sensitive to illumination changes and noises. The performance of LDP representation is evaluated with two machine learning methods: Template matching and Support Vector Machines (SVM) with different kernels. Here we going to use SVM which is a supervised classifier.

For Image classifier is used in medical diagnosis process for increase the accuracy on effected nuclei detection and segmentation. The surveyed for Cell Phase Identification using KNN classifier and Nuclei Tracking using matching process as in [9]. So this shows that, it produces the 94% accuracy of nuclei detection and tracking algorithm achieved more than 90% tracking accuracy. But Errors happen over segmentation problem when daughter cell nuclei overlap with nearby nuclei right after division and need to and improve the segmentation process. With the SVM classifier with 92% accuracy achievement as in [18].

In this Project, only Acute Myelogenous Leukemia (AML) is considered.This technique greatly depends on the operator's ability and fatigue levels.The most Common drawback observed from various survey from different literatureand many morewhich asa problem of over-segmentation of cluster of nucleus and accuracy of detection of effected nuclei problems. The problem of images with high degree of cell overlapping has overcome as in [18] with help of SVM classifier. Thus, themain goal of our paper is to implement a fully automated classifier system for AML and to increase the accuracy of finding exact effected cells and overcome the over-segmentation of cluster of nucleus problem which influence our goal.

III. SYSTEM METHODOLOGY

In this Section, the framework for the proposed segmentation method is described. The flow diagram of the proposed method is shown in Fig. 3 which consists of four stages: Pre-processing stage, Segmentation stage, feature extraction stage and Classification stage respectively. The details of the framework are given in the following subsections:

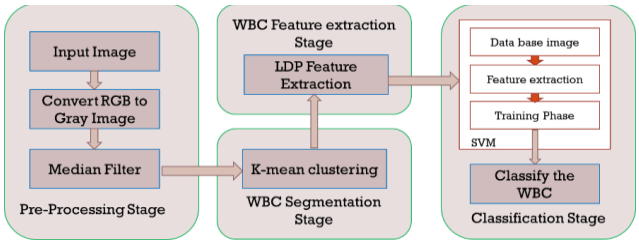


Fig. 3. Flow diagram of proposed system.

A. *Image Pre-Processing Stage*

In this stage, input image is obtained and image is resized as needed for further process. The resized image which is RBG color space converted into Gray scale level image. Finally, then theselective median filtering applied which used to remove unwanted effects such as noise from the image.

B. *Image segmentation stage*

In this stage, goal of image segmentation is to extract important information from an input image. It plays a key role since the efficiency of subsequent feature extraction and classification relies greatly on the correct identification of the myeloblasts.The Image segmentation has developed for gray-level images.Segmentation in this system is performed for extracting the nuclei of the leukocytes using K-mean clustering. K-means is still widely used as in [1], [13].In this paper, we chose clusters corresponding to only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity (background) to white at the strongest intensity (nucleus).Here, every pixel is assigned to one of these classes using the properties of the cluster center.

C. *Image Feature Extraction stage*

Feature extraction in image processing is a technique of redefining a large set of redundant data into a set of features of reduced dimension. Transforming the input data into the set of features is called feature extraction. Feature selection greatly influences the classifier performance; therefore, a correct choice of features is a very crucial step as in [17]. We implemented these features on whole images in our system. Those features were considered to boost the classifier performance.

Local Directional Pattern (LDP): LDP is a gray-scale texture pattern which characterizes the spatial structure of a local image texture. A LDP operator computes the edge response values in all eight directions at each pixel position and generates a code from the relative strength magnitude. Since the edge responses are more illumination and noise insensitive than intensity values, the resultant LDP feature describes the local primitives including different types of curves, corners, and junctions, more stably and retains more information. LDP operator use the edge response values of neighborhood pixels and encode the image texture. The LDP is computed as in [14].

The LDP assigns neighbor pixel, N as 8 bit binary code to each center pixel of an input image with window size, n as 1 andk=3 as 3x3 matrix. The Features information is extracted by histogram plot. To construct a histogram, the first step is to find Nbin value which is the range of values, i.e., that is, divide the entire range of values into a series of small intervals and then count how many values fall into each interval using (1).

$$\text{Nbin} = \frac{N!}{N! * (N-n)!} \quad (1)$$

The histogram is constructed between the number of occurrences of each pixel value and minimum and maximum pixel range. Then variation between normal and abnormal is predicted easily on histogram plot.

D. *Image classification stage*

The selection of a classification technique for classification is a challenging problem because an appropriate choice given the available data can significantly help improving the accuracy in credit scoring practice as in [1], [2]. There is a plenty of statistical techniques, which aim at solving binary classification tasks. In this project, we use a support vector machine (SVM) for constructing a decision surface in the feature space that bisects the two categories, i.e., cancerous and noncancerous, and maximizes the margin of separation between two classes of points. SVMs is a promising nonlinear nonparametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting, and other fields as discussed on [20], [23].

An SVM is primarily a two-class classifier. It can be either linear or nonlinear. In this paper, we choose a linear SVM two-class classifier; because it is not computationally expensive, it does not employ the kernel trick explicitly, and it achieves, in general, a good performance as in [20], [23] and [26]. The classification using the SVM, a statistical method called cross-validation is used for evaluating and comparing learning algorithms. Cross-validation is a technique for judging how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction and where one wants to estimate how precisely a predictive model will perform in practice. Three kinds of validation techniques have been used: *K*-fold, holdout, and leave-one-out as in [1], [2]. A quantitative analysis of the performance of the proposed method for segmentation is carried out in terms of accuracy. The accuracy is calculated from a dataset of sample images which are tested individually.

The SVM Classifier gets the normalized value of the histogram from feature extraction stage and thenSVM is trained with obtainedhistogram features. A collection of each sampleimage is tested. Then these tested image relate to the possible outcomes of the classifier system. When attempting to classify a specimen, there are four possibilities: True Positives (TP), when cancer cells are correctly identified; False Positives (FP), when non-cancer cells are identified as cancerous; True Negatives (TN), when non-cancerous cells are correctly identified; and False Negatives (FN), when cancer cells are identified as noncancerous as in [2]. Here, we classified as normal or abnormal conditional of the microscopy input image sample.These parameters are calculated, the overall accuracy of tested Image which is formulated as mentioned as shown in Table I given in [2], that higher values of these parameters will lead to better accuracy.

TABLE I. PARAMETER FOR PERFORMANCE EVALUATION

Parameter	Formule
Precision	P=(TP/(TP+FP))
Specificity	Sp=(TN/(TN+FP))
Sensitivity	Se=(TP/(TP+FN))
F-measure	F=(2*Sp*P)/(P+Se)
Accuracy	A=((Tp+Tn)/(Tp+Tn+Fp+Fn))

IV. COMPUTER SIMULATION RESULTS

The proposed technique has been applied on peripheral blood smear images obtained from diagnosis center, as aforementioned. A microscopic blood image of any size is considered for evaluation. The superiority of the scheme is demonstrated with the help ofsome sample of inputs.In this project, we can use MATLAB 7.5 and above version for analysis of diagnosis samples.

Pre-Processing Stage: Using MATLAB software, get the input blood sample image from data sample set. Let be, input image, Obtained Image is processed by gray scale conversion to get intensity information and then if any noise, then gray level image apply to median filter as shown in Fig. 4 (a)-(f),

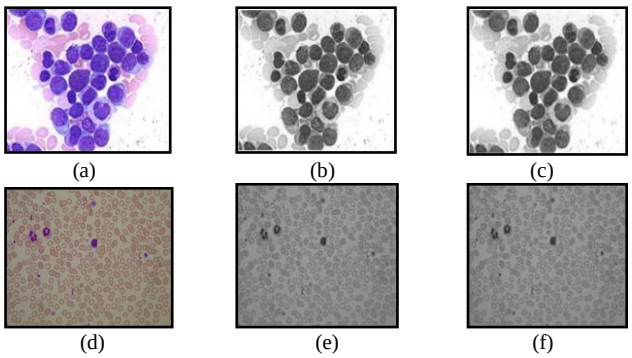


Fig. 4. (a), (d)Input sample image. (b), (e)Converted Gray scale image.and (c), (f)Filtered Gray scale image.

Image Segmentation Stage: Result obtained from Pre-processing stage will provide Gray level Intensity information of image for easier segmentation of WBC Nucleus from cluster of nuclei. This image segmentation done by using k-mean Clustering technique. This k-mean Clustering Technique to group your objects based on features into K number of group. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data. The Clusters of objects (k=4) for input image is shown in Fig. 5 for input image of Fig. 2 (a) and (b).Now, segmented image output for Fig. 2 (a) and (b) is shown in Fig. 6 (a) and (b).

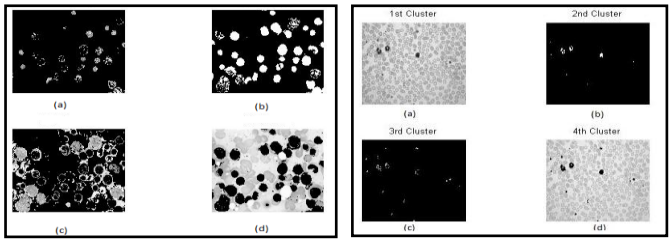


Fig. 5. Clustered of objectfor Fig. 2 (a) and (b). (a) 1st Cluster, (b) 2nd Cluster, (c) 3rd Cluster and (d) 4th Cluster

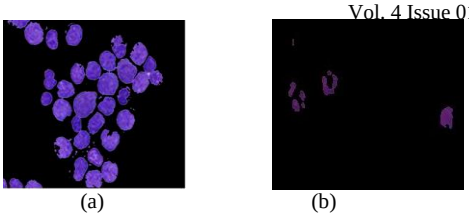


Fig. 6. (a) Final Nuclei Segmented Output image for Fig. 2 (a). (b) Final Nuclei Segmented Output image for Fig. 2 (b).

Image feature extraction stage: Nuclei is segmented from the previous stage and then LDP operator is applied for extracting features of segmented nuclei. Fig. 7 shows the texture analysis of LDP for Fig. 6 which as segmented image. The features of segmented image are obtainedby constructinga histogram plot. To construct histogram, we havecalculated the Number of bins value. Thus, small variation of number of count values in each pixelsis used to differentiate normal or abnormal case. The histogram plots of both normal and abnormal is shown in Fig. 8.

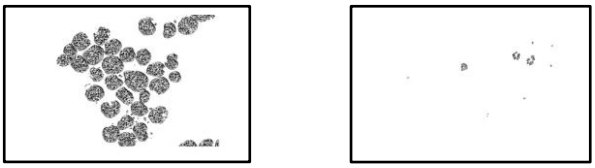


Fig. 7. LDP based Texture analysis of (a) nuclei segmented image of Fig. 6 (a) and (b) nuclei segmented image of Fig. 6 (b).

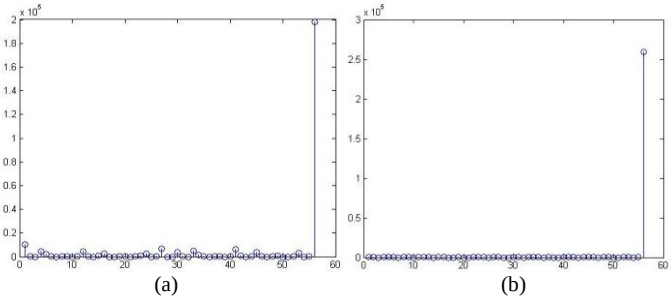


Fig. 8. Histogram plot for (a) Fig. 7 (a) and (b) Fig. 7 (b).

*Image classification stage:*The values of Histogram will be very high in range of thousands and this high range value can't be implemented in SVM classifier directly. So that normalized values are calculated and applied to SVM classifier. SVM is used to differentiate two status of patient i.e., whether Abnormal or Normal case for given blood sample under test. The proposed system result shows dialog box as normal orabnormal condition of patient. In Abnormal case, as mentioned earlier, larger number of WBC (cluster of Blue Nuclei-WBC) leads to the Leukemia cancer which shows abnormally of patient as shown in Fig. 9 (a) and large variations in number of count of each pixels of abnormal case as shown in Fig. 8 (a). Whereas, in Normal case, Count of WBC (Blue Nuclei) will less which shows patient does not has cancer which is shown in Fig. 9 (b)and small or no variations in number of count of each pixels of normal case as shown in Fig. 8 (b). The performance evaluation is calculated using the parameter values shown in Table II form tested images. The

Overall accuracy of our system compared with few existing system is shown in Fig. 10.

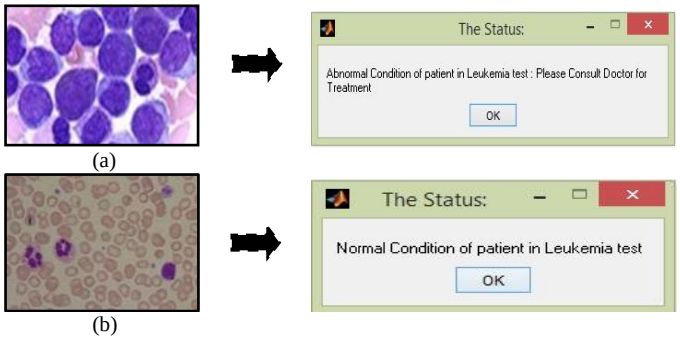


Fig. 9. (a) Abnormal condition and its Status. (b) Normal condition and its Status.

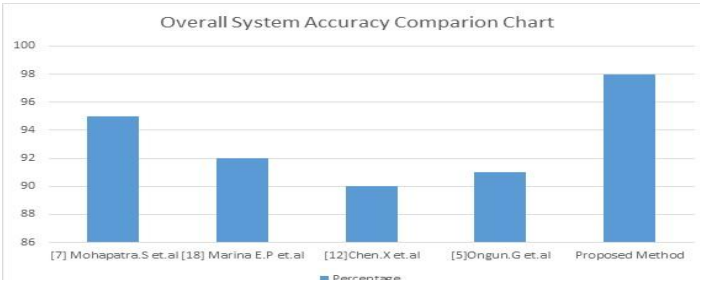


Fig.10. Overall accuracy of our system compared with few existing system.

TABLE II. PERFORMANCE EVALUATION TABLE

Parameters	Values
Sensitivity (%)	100
Specificity (%)	80
Precision (%)	85.7143
F-Measure (%)	93.4483

V. CONCLUSION

This proposed system has been successfully the implemented and evaluatedfor an automated screening system for AML in blood microscopic images. It uses 90 high-quality 184 × 138 size images obtained from the American Society of Hematology, which has to be tested for leukemia. A feature set exploiting the texture factor of a blood cell image to get all the information required using histogram to do efficient classification. The impact of the LDP operator on the segmented image provided a promising feature for this analysis. Thus, 88 out of 90 images yields correct results of 98%, which shows the accuracy of tested images of the proposed system and presents a good demarcation between cancer and non-cancer status of a patient under test.

Further, research will focus on the collection of more samples to yield better performance and building an overall system for cancer classification. Then in future this diagnostic method can be tested for various diseases such as Ebola, Tumor, Brain cancer, etc.

ACKNOWLEDGMENT

The authors would like to thank the American Society of Haematology and MedlinePlus for providing a high-quality image database.

REFERENCES

[1] D. Mandal, K. Panetta, and S. Agaian, "Face recognition based on logarithmic local binary patterns," in *Proc. SPIE, Image Process., AlgorithmsSyst. XI*, 2013, vol. 8655, pp. 865514-1–865514-12.

[2] Agaian S., Madhukar M., and Chronopoulos A.T. (2014), "Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images", IEEE Systems Journal, Vol. 8, No. 3.

[3] Haworth C., Palmer Mk., Heppleston Ad., Jones M. Ph., Rha C., Dik E. (1981), "Routine bone inarrow examination in the management of acute lymphoblastic leukaemia of childhood", J Clin Pathol; Vol. 34, pp. 483-485.

[4] MedlinePlus: Leukemia.National Institutes of Health. [Online]. Available:http://www.nlm.nih.gov/medlineplus/ency/article/001299.htm

[5] F. Scotti, "Automatic morphological analysis for acute leukemia identi-fication in peripheral blood microscope images," in *Proc. CIMSA*, 2005, 96–101.

[6] ASH Image Bank: American Society of Hematology. [Online]. Available: http://imagebank.hematology.org/Default.aspx

[7] A. Nasir, M. Mashor, and H. Rosline, "Unsupervised colour segmentation of white blood cell for Acute leukaemia images," in Proc. IEEE IST, 2011, pp. 142–145.

[8] Nallaperumal K. and Krishnaveni K. (2008), "Watershed segmentation of cervical images using multiscale morphological gradient and HSI color space," Int. J. Imaging Sci. Eng., Vol. 2, No. 2, pp. 212–216.

[9] Chen X., Zhou X., and Stephen T.C. W. (2006), "Automated Segmentation, Classification, and Tracking of Cancer Cell Nuclei in Time-Lapse Microscopy", IEEE Transactions on Biomedical Engineering, Vol. 53, No. 4.

[10] Rangayyan R. (2004), "Biomedical Image Analysis", Biomedical Engineering. Boca Raton, FL, USA: CRC Press.

[11] Ilea D. and Whelan P. (2011), "Image segmentation based on the integration of colour-texture descriptors—A review," Pattern Recognit., Vol. 44, No. 10/11, pp. 2479-2501.

[12] S. Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," in Proc. IECR, 2010, pp. 215–219.

[13] Ravichandran K. S. and Ananthi B. (2009), Color Skin Segmentation using K–Means Cluster. International Journal of Computational and Applied Mathematics, Vol. 4, No.2, pp.153 – 157.

[14] Jabid T., Kabir M. H. and Chae O. S. (2010), "Local Directional Pattern (LDP) for Face Recognition", Proceeding of the IEEE International Conference of Consumer Electronics.

[15] Adollah R., Mashor M., Nasir N., Rosline H., Mahsin H., and Adilah H. (2008) "Blood cell image segmentation: A review," in Proc. IFMBE. Berlin, Germany: Springer-Verlag, Ch. 39, pp. 141–144.

[16] Marina E. P., Christophoros Nikou., and Antonia Charchanti. (2011), "Automated Detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering", IEEE Transactions on Information Technology in Biomedicine, Vol. 15, No. 2.

[17] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system," in *Proc. IJCNN*, 2001, vol. 4, 2461–2466.

[18] Markiewicz T., Osowski S., Marianska B., Moszczynski L., (2005), "Automatic Recognition of the Blood Cells of Myelogenous Leukemia Using SVM", Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, pp. 2496-2501.

[19] C. Lopez and S. Agaian, "Iterative local color normalization using fuzzy image clustering," in *Proc. SPIE, Mobile Multim./Image Process., Security,Appl.*, 2013, vol. 8755, pp. 875518-1–875518-12.

[20] Ojala T. and Pietikainen M. (2002), "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, pp. 971-987.

[21] C. Lopez and S. Agaian, "A new set of wavelet- and fractals-based features for Gleason grading of prostate cancer histopathology images,"

in *Proc. SPIE, Image Process., Algorithms Syst. XI*, 2013, vol. 8655, pp. 865516-1–865516-12.

[22] Agarwal A., Chapelle O., Dudik M. and Langford J. (2011), "A reliable effective terascale linear learning system," arXiv: 1110.4198, preprint, Cornell University Library.

[23] H. Inaba, M. Greaves, and C. Mullighan, "Acute lymphoblastic leukaemia," Lancet, vol. 381, no. 9881, pp. 1943–1955, Jun. 2013.

[24] J. Tang and S. Agaian, "Computer-aided cancer detection and diagnosis: Recent advances," in *Proc. SPIE*. Bellingham, WA, USA: SPIE, 2014,p. 341.

[25] A. Almuntashri, E. Finol, and S. Agaian, "Automatic lumen segmentation in CT and PC-MR images of abdominal aortic aneurysm," in *Proc. IEEEInt. Conf. Syst., Man Cybern.*, 2012, pp. 2891–2896.