Comparison of Segmentation Algorithms for Leukemia Classification

Sunita Chand, Virendra P. Vishwakarma
[1]University School of Information and Communication Technology, Guru Gobind Singh University, New Delhi, India
[1]sunitamk@gmail.com
[2]University School of Information and Communication Technology, Guru Gobind Singh University, New Delhi, India
[2]virendravishwa@rediffmail.com

Abstract: Leukemia is a deadly cancer that results from the proliferation of non-differentiated white blood cells in blood as compared to the other two types of cells, i.e., red blood cells and platelets. These cells are known as blasts cells which overcrowd other cells rendering those cells as inefficient in their functions and are are themselves non-functional. This paper presents a comparative study of four different segmentation techniques on the images of peripheral blood smear and the classification of these images into diseased and healthy cells using the SVM classifier. The best result was obtained by a custom threshold method of segmentation with a classification accuracy of 96.89%.

Keywords: Acute Leukemia, Machine Learning, Support Vector Machine. Image Processing, Image Segmentation

1 Introduction

Human Blood mainly consists of three types of cells, i.e., white blood cells(WBC), red blood cells(RBC) and platelets. RBCs play an essential role in the transportation of oxygen to all body organs and the removal of carbon dioxide through the blood. White blood cells, on the other hand, are the fighter cells of our body which help our body to fight infections, whereas platelets help in blood coagulation[1]–[3][4]. All these types of cells are developed from stem cells produced in the bone marrow present in bones, via a process known as hematopoiesis. Leukemia, also known as 'liquid cancer' or 'cancer of blood' may result due to exposure to radiations, certain chemicals or due to genetic history[5]. As per American Cancer Society statistics, 60,530 new leukemia cases are estimated in 2020 and estimated deaths are projected to be 23,100. Leukemia is the seventh leading cause of cancer death in USA as per the National Cancer Institute. It is preliminarily diagnosed by the complete blood count (CBC) clinical blood test and is caused by the uncontrolled accumulation of partially developed leukocytes or white blood cells, also known as blast cells. Based on the progression of disease, Leukemia may be categorized as acute leukemia, characterized by fast progression and speedy deterioration of patient and chronic leukemia, characterized by slow progression of disease. All components of blood originate either from myeloid progenitor cells or lymphoid progenitor cells. Hence Leukemia may be further categorized on the basis of origin of lymphoblasts and the speed of progression into various categories as shown in figure-1. Though the disease is diagnosed by a preliminary assessment of blood components, the final diagnosis is dependent on biopsy of bone marrow, cytogenetics, immunophenotyping, or PET scan tests. Analyzing blood smear or bone marrow biopsy under the microscope is very time consuming, and the results chiefly depend on the expertise of the hematopathologists.[6],[7], which calls the need to automate the process of diagnosis[8][9]. Various researchers from the field of medicine, computer science, and hematopathologists are working together to develop a sound model of leukemia diagnosis[10].

ICASISET 2020, May 16-17, Chennai, India

DOI 10.4108/eai.16-5-2020.2303967

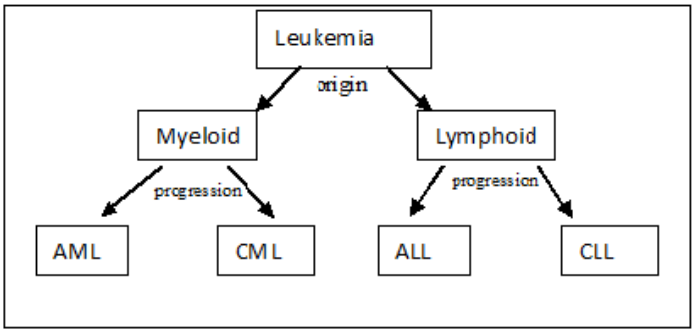

Figure 1: Classification of Leukemia

Image processing techniques and machine learning have proved to be a boon for this task. Researchers have tried and tested various methods of classification to the features extracted from the images using image processing techniques[11][12]. The images of blood smears are first captured, preprocessed for various types of enhancements, and segmented to find the region of interest(ROI)[13]–[15][16]. The features extraction and selection are then carried out, which are then fed into machine learning-based classifiers for classification into healthy or diseased. The classification can also be applied to identify different types of leukemia, i.e., acute or chronic, AML or ALL, or further categories of AML or ALL, etc.[17], [18][19]–[21].Various machine learning techniques have been used for classification in the literature e.g., SVM[8], [22]–[29], KNN[30], [31], decision tree[6], [14], [21], [32]–[35] and random forests[32], [36]–[40] to name a few.

The paper is further described in various following sections. Section 2 gives an overview of the proposed model. Section 3 provides the segmentation techniques used; section 3 gives a brief of the classification method used, and section 4 describes the results and its analysis. Finally, the concluding remarks are given in section 5.

The proposed Model

The paper presents the study of four different segmentations on classification, as shown in figure 2.

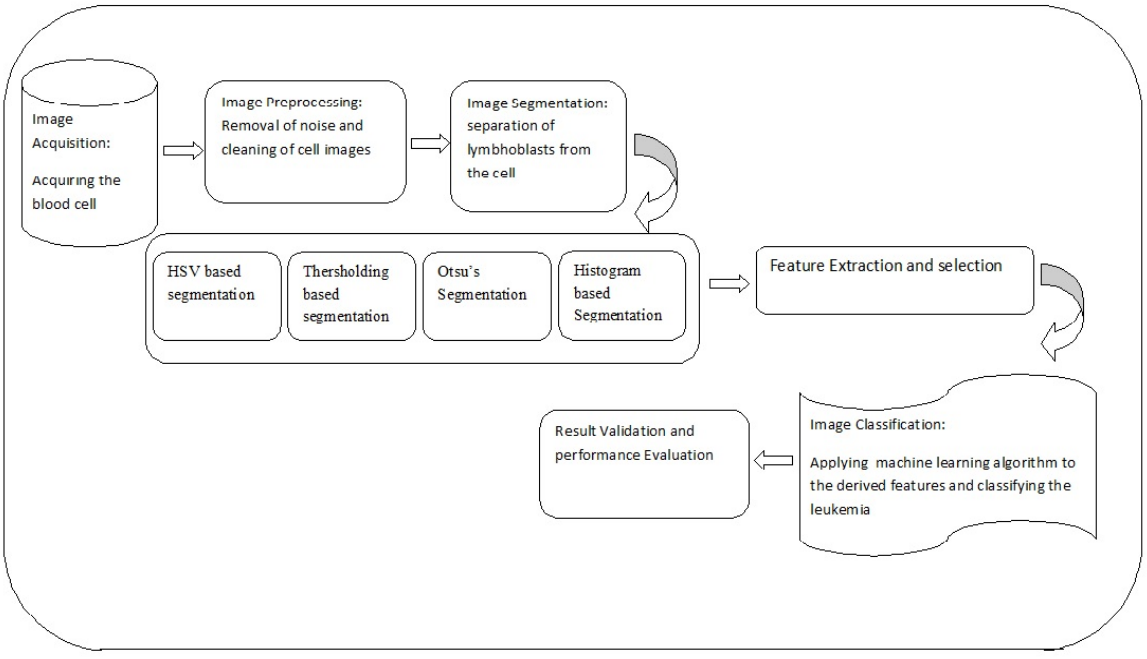

Figure 2: Research Framework

As described in the literature, image processing techniques include image acquisition, image enhancement, image segmentation, feature extraction, and classification.

Step 1: Acquiring Dataset

The study of leukemia has been affected by the lack of availability of public datasets [14], [28][41][13], [15]. The only dataset which is in the public domain and is available on request is ALL-IDB from Fabio Scotti,R.D. Labatti, and Vincenzo Piuri, which has been extensively used by the researchers in this domain.[[42]–[44]. The dataset consists of two datasets ALL-IDB1 and ALL-IDB2. ALL-IDB1 consists of 108 blood smear images collected during September 2005. It comprises of 49 ALL affected images from leukemic patients while 59 images from healthy people. On the other hand, ALL-IDB2 consists of the segmented area from both normal and diseased cell images from the ALL-IDB1 dataset and consists of 260 images of segmented leukocytes or lymphoblasts having fifty percent each [47][48].

Step 2: Image Enhancements

Images are enhanced to remove the noise and get better results. Various types are enhancements are done to ALL-IDB 1 images like RGB to gray conversion, histogram stretching, separating HSV components, etc.

Step 3: Image segmentation

The enhanced images are then segmented using:

a) **Histogram stretching based segmentation**
b) **Manual Thresholding**
c) **HSV based Segmentation**
d) **Otsu's Segmentation**

a). Histogram based segmentation

Preprocessing :

i) Image converted to greyscale (E1)
ii) Image enhancement is performed using contrast stretching (E2)
iii) Histogram equalization is performed on contrast stretched image(E3)

Segmentation:

i) Added image (S1) = E2 + E3
ii) Subtracted Image (S2) = E2 –E3
iii) Background Removed Image = S1 + S2

The result of histogram-based segmentation is given below in figure 3, and figure 4.The result of manual thresholding, which was set at 80, is depicted in figure 5.The result of automatic Otsu's segmentation is illustrated in figure 6, while figure 7 shows the results of HSV based segmentation.

b) Manual Thresholding

Pixel intensity values were observed with Matlab inbuilt tool, and then several threshold values were tried and tested out of which pixels in the image with intensity greater than 80 were assigned a pixel value equal to 1.

c) HSV based segmentation

Image is converted from RGB(Red,Green,Blue) to HSV color model wherein H stands for Hue(color) component, S for Saturation, i.e., shade or amount of gray,and V stands for brightness value. S component of HSV is extracted from the image as it gives the structural information of the leucocyte nucleus[45], [46].

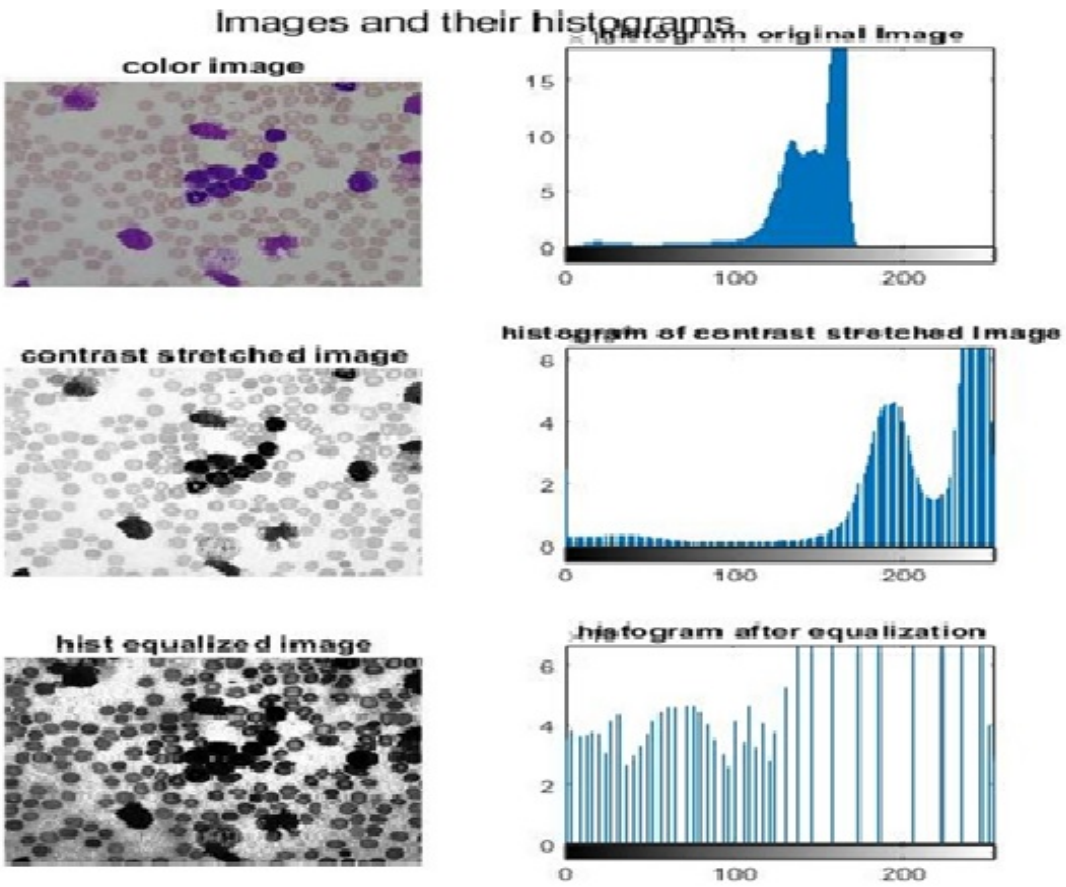

Figure 3: Histogram of images

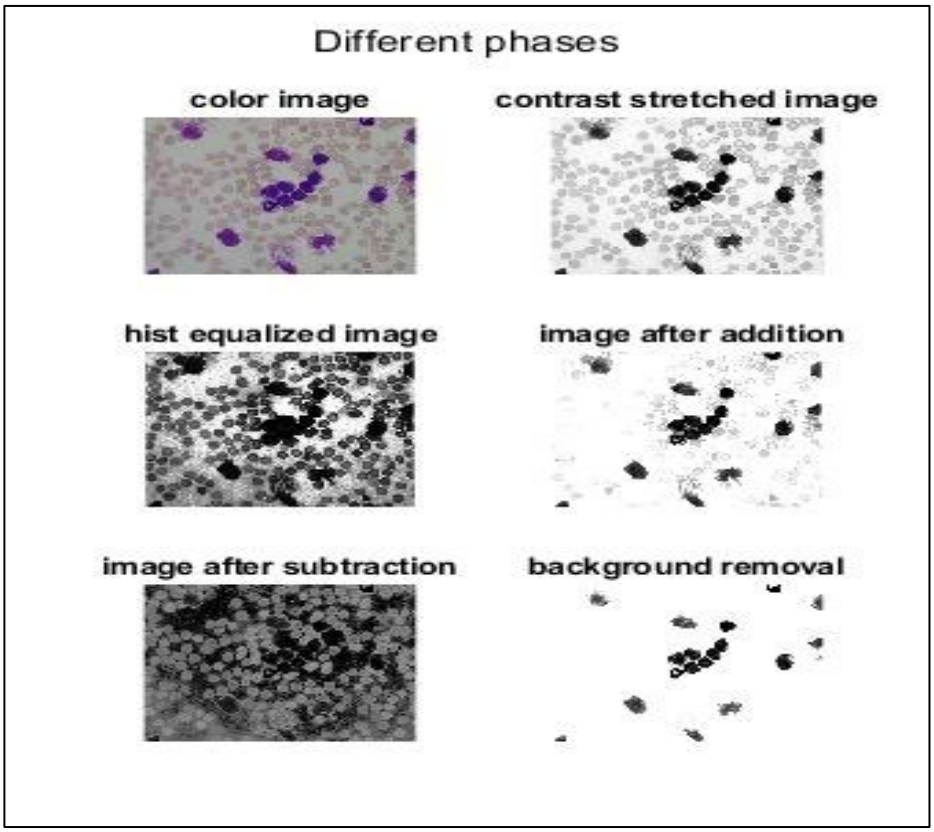

Figure 4: Histogram based Segmentation

d) Otsu's Segmentation

The Otsu's thresholding method proceeds by iterating through all the possible threshold values, and measures the spread of the pixel levels along each side of the threshold to find whether the pixels fall in foreground or background. A threshold value having the minimum value for the sum of the foreground and the background spreads is thus found.

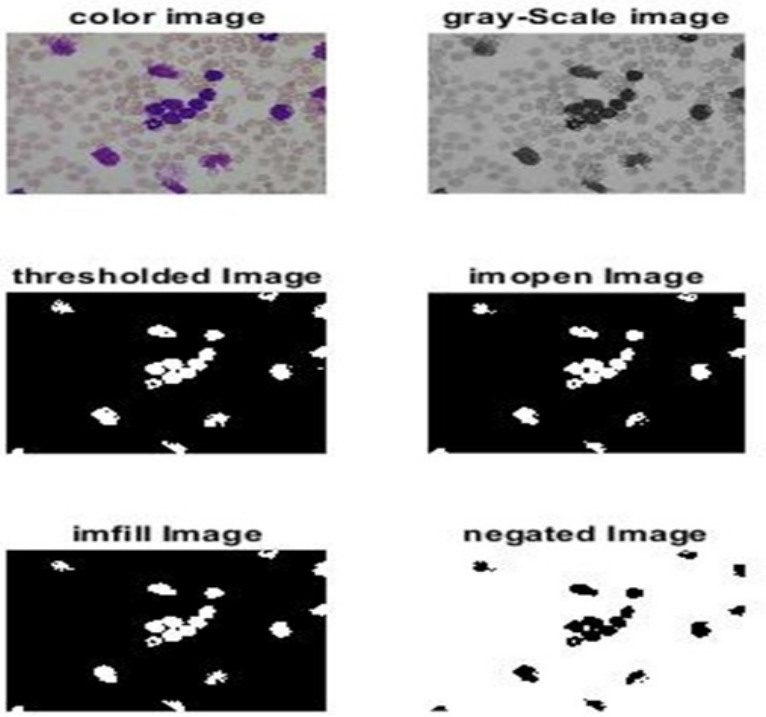

Figure 5: myThresh Segmentation

Step: 4 Feature Extraction:

After successful segmentation, the connected components are found, and for each connected component, the following morphological features are extracted from each blood cell image. The following features are then averaged to find the mean features which are then fed into the SVM classifier.

Area	perimeters	majorAxis	minorAxis	eccentricities
orientations	convexAreas	FilledArea	Eq_Diameter	solidity
extent	Roundness			

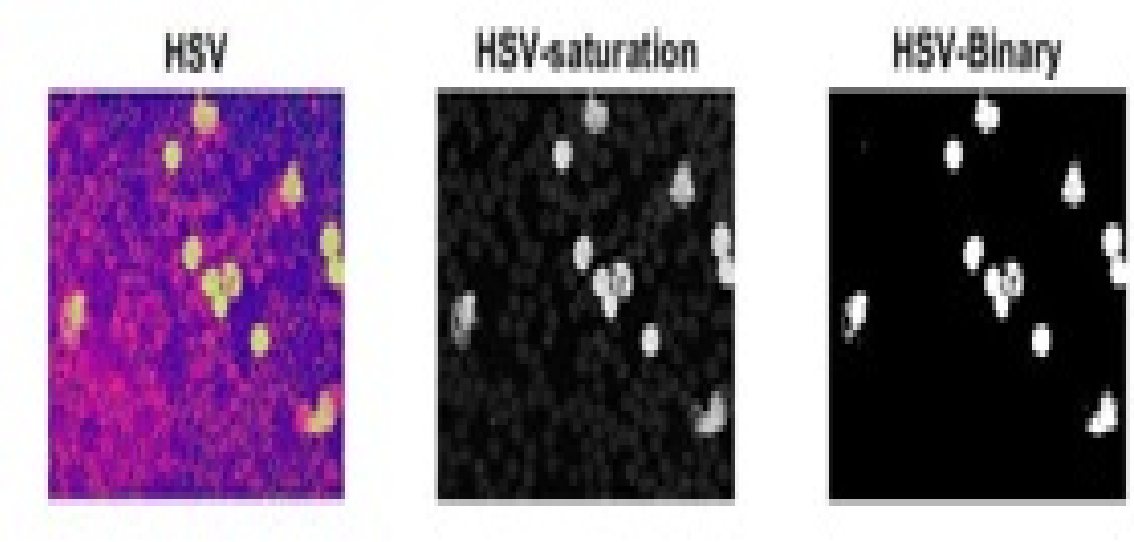

Figure 6: HSV based segmentation

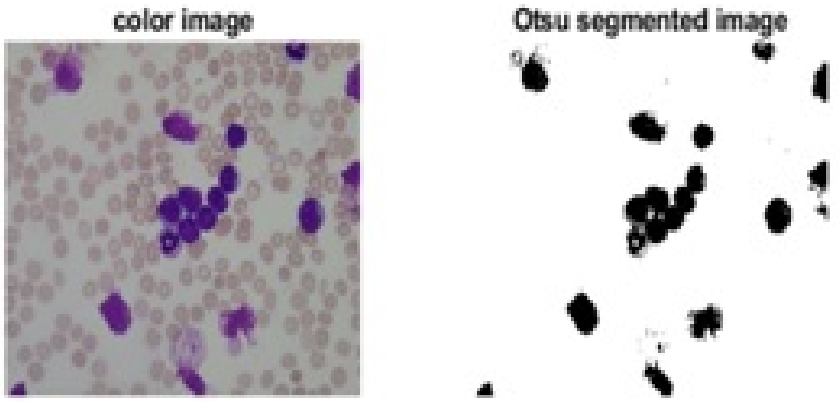

Figure 7: Otsu's Segmentation

3. Classification

The mean features thus obtained are further fed into the SVM classifier. We implemented the classification using a two-class SVM classifier with a radial basis kernel. SVM has been extensively used by researchers to classify the blood smear images. The input images are divided into training image data and testing image data in various ratios, i.e. (70:30), (80:20), and (75:25), etc. The SVM model is trained using the training data. K-fold cross-validation is used with k=10 to avoid over-fitting and over-training.

In this method, the training data is divided into K equal-sized folds or sets. In each of the K-iteration, one set or fold is used for validation, and rest K-1 folds are used for training, as shown in figure 8.

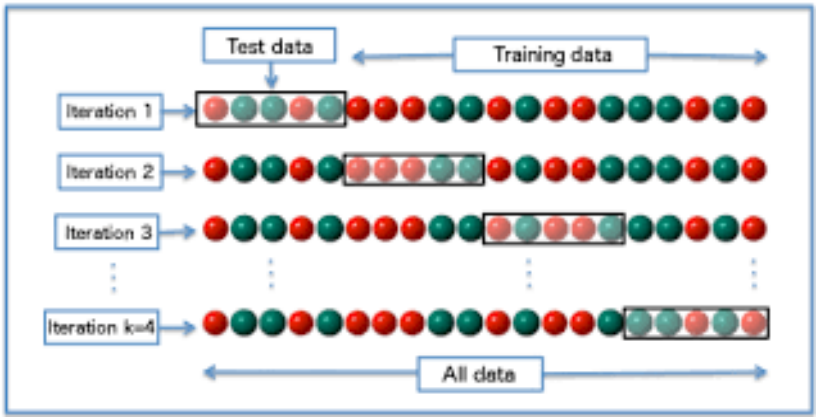

Figure 8: Cross-Validation

Image courtesy- https://en.wikipedia.org/wiki/Cross-validation_(statistics)

4. Results and Discussion:

The SVM classification results for various segmentation algorithms are summarized in table 1 given below:

Table 1: SVM Classification Results

S.No	Segmentation	Accuracy
1	HSV	88.9%
2	MyThreshold	96.89%
3	Otsu	86.8%
4	Hist_Contrast	86.95%

The results demonstrate that the same SVM gives different results when applied to the images segmented by various techniques. Thus it directly supports the effectiveness of segmentation methods and its correctness. The better the segmentation, the better will be the classification. The experiment was conducted with a small set of segmentation methods.
The same experiment may be performed with a broader set of segmentation techniques. The given research demonstrates that the feature extraction and selection also depend on the correctness of segmentation. It was found that the threshold of 80, which was manually selected to divide the intensity values, gives the best result among other segmentation i.e. 96.89%

References

[1] T. S. Ahmed, K. Philip, and A. E. Ahmed, "Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features," *Sci. Rep.*, pp. 1–11, 2020.

[2] S. Chin Neoh *et al.*, "An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images," *Sci. Rep.*, vol. 5, pp. 1–14, 2015.

[3] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukemia detection using image processing techniques," *Artif. Intell. Med.*, vol. 62, no. 3, pp. 179–191, 2014.

[4] M. A. Alsalem *et al.*, "A review of the automated detection and classification of acute leukaemia: Coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations," *Comput. Methods Programs Biomed.*, vol. 158, pp. 93–112, 2018.

[5] T. T. P. Thanh, C. Vununu, S. Atoev, S. Lee, and K. Kwon, "Leukemia Blood Cell Image Classification Using Convolutional Neural Network," *Int. J. Comput. Theory Eng.*, vol. 10, no. 2, 2018.

[6] K. A. S. A. Daqqa and A. Y. A. Maghari, "Classification Algorithms," pp. 638–643, 2017.

[7] B. . devi C. Vidhya, P.S Kumar, K.Keerthika, C.Nagalakshmi, "Classification of Acute Lymphoblastic Leukemia in Blood Microscopic Images using SVM," *Int. Conf. Eng. Trends Sci. Humanit.*, vol. 4, no. 01, pp. 232–236, 2015.

[8] O. Sarrafzadeh, H. Rabbani, A. Talebi, and H. U. Banaem, "Selection of the best features for leukocytes classification in blood smear microscopic images," no. July, p. 90410P, 2014.

[9] J. Rodellar, S. Alférez, A. Acevedo, A. Molina, and A. Merino, "Image processing and machine learning in the morphological analysis of blood cells," *Int. J. Lab. Hematol.*, vol. 40, no. February, pp. 46–53, 2018.

[10] H. Abedy, F. Ahmed, and N. Q. Bhuiyan, "Leukemia Prediction from Microscopic Images of Human Blood Cell Using HOG Feature Descriptor and Logistic Regression," *2018 Sixt. Int. Conf. ICT Knowl. Eng. Leuk.*, 2018.

[11] S. Agaian, M. Madhukar, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," *IEEE Syst. J.*, vol. 8, no. 3, pp. 995–1004, 2014.

[12] F. B. Jada, A. M. Aibinu, and A. J. Onumanyi, "Performance Metrics for Image Segmentation Techniques : A Review," no. October, pp. 344–348, 2015.

[13] H. B. Kekre, D. Ph, and P. D. Asst, "Segmentation of Blast using Vector Quantization Technique," *Int. J. Comput. Appl.*, vol. 72, no. 15, pp. 20–23, 2013.

[14] A. S. Negm, O. A. Hassan, and A. H. Kandil, "A decision support system for Acute Leukaemia classification based on digital microscopic images," *Alexandria Eng. J.*, 2017.

[15] P. Viswanathan, "Fuzzy C Means Detection of Leukemia Based on Morphological Contour Segmentation," *Procedia Comput. Sci.*, vol. 58, pp. 84–90, 2015.

[16] S. C. ; V. P. Vishwakarma, "Leukemia Diagnosis using Computational Intelligence," in *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),GHAZIABAD, India.*, 2019, pp. 1–7.

[17] A. Setiawan, A. Harjoko, T. Ratnaningsih, E. Suryani, and S. Palgunadi, "Classification of Cell Types In Acute Myeloid Support Vector Machine Classifier," *Int. Conf. Inf. Commun. Technol.*, no. Cml, pp. 45–49, 2018.

[18] A. Setiawan, A. Harjoko, T. Ratnaningsih, E. Suryani, Wiharto, and S. Palgunadi, "Classification of cell types in Acute Myeloid Leukemia (AML) of M4, M5 and M7 subtypes with support vector machine classifier," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, 2018, vol. 2018-January.

[19] A. Harjoko, T. Ratnaningsih, E. Suryani, Wiharto, S. Palgunadi, and N. P. T. Prakisya, "Classification of acute myeloid leukemia subtypes M1, M2 and M3 using active contour without edge segmentation and momentum backpropagation artificial neural network," *MATEC Web Conf.*, vol. 154, p. 01041, 2018.

[20] G. Sumalatha and N. J. R. Muniraj, "Survey on medical diagnosis using data mining techniques," *2013 Int. Conf. Opt. Imaging Sens. Secur. ICOSS 2013*, pp. 1–8, 2013.

[21] B. U. Maheswari and V. S. Joy, "Machine Learning Approach for Identifying and Diagnosing the Cancer Types," *Int J Adv Engg Tech/Vol. VII/Issue I/Jan.-March*, vol. 769, p. 773, 2016.

[22] S. Rajpurohit, S. Patil, and N. Choudhary, "Microscopic Blood Image Using Image Processing," *2018 Int. Conf. Adv. Comput. Commun. Informatics*, no. Cll, pp. 2359–2363, 2018.

[23] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Inform.*, vol. 2, pp. 59–77, 2006.

[24] L. Wang, L. Zhang, S. Member, M. Zhu, and Z. Yi, "Automatic Diagnosis for Thyroid Nodules in Ultrasound Images by Deep Neural Networks," *Med. Image*

Anal., p. 101665, 2019.
[25] A. I. Shahin, Y. Guo, K. M. Amin, and A. A. Sharawi, "White blood cells identification system based on convolutional deep neural learning networks," *Comput. Methods Programs Biomed.*, vol. 168, pp. 69–80, 2019.
[26] M. A. R. J. Begum and T. A. Razak, "DIAGONISING LEUKEMIA FROM MICROSCOPIC IMAGES USING IMAGE ANALYSIS AND PROCESSING TECHNIQUES," 2016.
[27] K. R. T. A. Luis H.S. Vogado *, Rodrigo M.S. Veras, Flavio. H.D. Araujo, Romuere R.V. Silva, "Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification," *J. Contin. High. Educ.*, vol. 60, no. 2, pp. 93–109, 2012.
[28] J. Rawat, A. Singh, B. HS, J. Virmani, and J. S. Devgun, "Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia," *Biocybern. Biomed. Eng.*, vol. 37, no. 4, 2017.
[29] S. Shafique and S. Tehsin, "Computer-Aided Diagnosis of Acute Lymphoblastic Leukaemia," *Comput. Math. Methods Med.*, vol. 2018, 2018.
[30] A. Rehman, N. Abbas, T. Saba, S. I. ur Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," *Microsc. Res. Tech.*, vol. 81, no. 11, pp. 1310–1317, 2018.
[31] S. Kumar, S. Mishra, P. Asthana, and Pragya, "Automated detection of acute leukemia using K-mean clustering algorithm," in *Advances in Intelligent Systems and Computing*, 2018, vol. 554.
[32] M. N. Q. Bhuiyan, S. K. Rahut, R. A. Tanvir, and S. Ripon, "Automatic acute lymphoblastic leukemia detection and comparative analysis from images," *2019 6th Int. Conf. Control. Decis. Inf. Technol. CoDIT 2019*, pp. 1144–1149, 2019.
[33] E. M. F. El Houby, "Framework of Computer Aided Diagnosis Systems for Cancer Classification Based on Medical Images," *J. Med. Syst.*, vol. 42, no. 8, 2018.
[34] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015.
[35] H. Shin and M. K. Markey, "A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples," *J. Biomed. Inform.*, vol. 39, no. 2, pp. 227–248, 2006.
[36] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, "Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection," *Biomed. Signal Process. Control*, vol. 33, pp. 272–280, 2017.
[37] S. Mishra, B. Majhi, and P. K. Sa, "Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection," *Biomed. Signal Process. Control*, vol. 47, pp. 303–311, 2019.
[38] L. H. S. Vogado, R. D. M. S. Veras, A. R. Andrade, F. H. D. De Araujo, R. R. V. E. Silva, and K. R. T. Aires, "Diagnosing Leukemia in Blood Smear Images Using an Ensemble of Classifiers and Pre-Trained Convolutional Neural Networks," *Proc. - 30th Conf. Graph. Patterns Images, SIBGRAPI 2017*, no. October, pp. 367–373, 2017.
[39] J. Cao and Z. Lin, "Extreme Learning Machines on High Dimensional and Large Data Applications: A Survey," *Math. Probl. Eng.*, vol. 2015, pp. 16–18, 2015.
[40] H. Mohamed *et al.*, "Automated Detection of White Blood Cells Cancer Diseases," *2018 First Int. Work. Deep Represent. Learn.*, pp. 48–54, 2018.
[41] M. B. Abdulkareem and S. Gaikwad, "Automatic Leukemia Identification System Using Otsu Image segmentation and MSER Approach for Microscopic Smear Image Database," *2018 Second Int. Conf. Inven. Commun. Comput. Technol.*, no. Icicct, pp. 267–272, 2018.
[42] F. S. Ruggero Donida Labati, Vincenzo Piuri, "ALL-IDB : THE ACUTE LYMPHOBLASTIC LEUKEMIA IMAGE DATABASE FOR IMAGE PROCESSING Ruggero Donida Labati , Vincenzo Piuri , Fabio Scotti Università degli Studi di Milano , Department of Information Technology ," *2011 18th IEEE Int. Conf. Image Process.*, pp. 2045–2048, 2011.
[43] F. Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images," no. July, pp. 20–22, 2005.
[44] V. Piuri and F. Scotti, "Morphological Classification of Blood Leucocytes by Microscope Images," no. July, pp. 14–16, 2004.
[45] P. Jagadev, "AUTOMATED DETECTION AND CLASSIFICATION OF LEUKEMIA USING IMAGE PROCESSING AND MACHINE LEARNING," *Int. J. Adv. Res. Sci. Eng.*, vol. 6, no. 4, pp. 324–329, 2017.
[46] P. Jagadev, "Detection of Leukemia and its Types using using Image Processing and Machine Learning," *Int. Conf. Trends Electron. Informatics ICEI 2017 Detect.*, vol. 1, no. 4, pp. 522–526, 2017.
[47] Vinod Jagannath Kadam, Shivaji rao Manik rao Jadhav, K.Vijayakumar, "Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Au

to encoders and Soft max Regression", Image & Signal Processing, springer, june 2019.

[48] K. Vijayakumar , K. Pradeep Mohan Kumar ,Daniel Jesline, "Implementation of Software Agents and Advanced AoA for Disease Data Analysis", journal of medical systems, Part of Springer Nature 2019.