

Detection of acute lymphoblastic leukemia using microscopic images of blood

Raju Bhukya *, B. Prasanth, V. Sasank Vihari, Y. Ajay

Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana, India

ARTICLE INFO

Article history:

Received 19 March 2017

Received in revised form

1 July 2017

Accepted 5 July 2017

Keywords:

Leukemia

Leukocyte

Microscopic image

Segmentation

Nucleus

Cytoplasm

ABSTRACT

Leukemia is a group of cancers that usually begin in the bone marrow and result in high number of abnormal white blood cells. Detection of leukaemia in early stages is necessary as this can reduce the rate of mortality and may lead to death. In a manual method of leukaemia detection Haematologists analyze the microscopic images and decide the severity. This is lengthy, cost effective and time taking process which depends on person's expertise and may not lead to standard accuracy. Till date, a number of methods have been proposed for this Leukaemia detection using Image Processing. Unlike the previous methods, which solely depend upon the entire cell, in this paper we proposed a new method to separate the cell Nucleus from Cytoplasm to obtain more features. The proposed method achieves the better accuracy when compared to the other existing methods.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Cancer can be defined as a disease in which a group of abnormal cells grow uncontrollably by disregarding the normal rules of cell division. There are several types of cancers which includes Skin cancer, Lung cancer, Prostate cancer, Blood cancer etc. Leukemia comes under blood cancer. Leukemia is a type of cancer pertaining to white blood cells (WBCs), whereby abnormal and immature WBCs are produced by the bone marrow and enter the blood (Turgeon, 2005). Generally, cells grow and multiply themselves to form new cells as the body needs them for effective functioning. When cells grow old, they die and new cells take their place. Sometimes, this cycle does not work properly (Majno and Joris, 2004). In cancer, new cells are formed when the body does not need them, and old cells do not die when they should (Patel and Mishra, 2015). Thus the number of abnormal white blood cells become numerous and interferes with normal white blood cells to carry out their functions. This also causes an imbalance of blood system in human body (Vaghela et al., 2015). The two main characteristics of cancer are uncontrolled growth of the abnormal cells in the human body and the ability of these cells to migrate from the original site and spread to distant sites. If

the spread is not controlled, cancer can result in death. There are two types of acute leukemia, namely acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) (Turgeon, 2005). Acute lymphoblastic Leukemia (ALL) is diagnosed in 3000 to 4000 persons in the United States each year; two thirds of them are children. The current rate of cure of nearly 80 percent in children attests to remarkable progress in the development of effective treatments for resistant subtypes of the disease (Pui and Evans, 1998). Acute leukemia is usually diagnosed by a morphological analysis of blood slides by hematologists, which is a complex, time-consuming, and costly process (Mohapatra et al., 2012a). It also requires considerable training and expertise.

Furthermore, the results often lack a standardized performance owing to a variety of factors including insufficient expertise or imperfection of the samples (Piuri and Scotti, 2004; Mohapatra and Patra, 2010). Unfortunately, the accurate assignment of patients to specific risk groups is a difficult and expensive process, requiring intensive laboratory studies including immune phenotyping, cytogenetics, and molecular diagnostics. Moreover, these diagnostic methods require the collective expertise of a number of professionals, and although this expertise is available at most major medical centers, it is generally unavailable in developing, under-developed countries (Yeoh et al., 2002). Microscopic examination of bone marrow blood smear and aspirate is the most economical technique but

* Corresponding Author.

Email Address: drrajunitw@gmail.com (R. Bhukya)

<https://doi.org/10.21833/ijaas.2017.08.011>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

manual examination of microscopic images (Srisukkham et al., 2013) often leads to bias because the results lack standardized performance owing to a variety of factors including insufficient expertise of the person, imperfection of the samples (Vaghela et al., 2015). Thus, there is a need to automate this procedure (Shirvoikar and Virani, 2016). Various digital diagnosis systems were developed to analyze microscopic blood images for leukemia detection. However, they suffered from a number of problems and limitations, in particular accurate diagnosis of leukemia require discrimination of one cell type from another, and of the cell nucleus from cell cytoplasm.

Indeed, separation of Leukemia cell nucleus with diverse complex irregular morphology from cytoplasm is a challenging task. Research shows that only a few existing clustering algorithms are able to achieve good adaptivity for reliable separation of nucleus and cytoplasm (Mohapatra et al., 2012a; Mohapatra et al., 2012b; Mohapatra et al., 2014). Therefore, the existing methods like k-means (Fatma and Sharma, 2014), FCS1 (Li et al., 2011), FCS2 (Wu et al., 2005), FCM (Bezdek et al., 1984), LDA (Li et al., 2011) are not reliable because of the limitations of the clustering algorithms (Kuo and Landgrebe, 2004).

This paper aims to overcome the above mentioned challenges. Here we propose a new segmentation algorithm to efficiently separate cell nucleus from its cytoplasm. The proposed algorithm uses discriminant measures to separate nucleus and cell where it uses features representing shape and color extracted from the information based on cell, cytoplasm and nucleus. The employed classifier, Support vector machine (SVM) take the statistical values obtained from the features to classify healthy and unhealthy cells (Tong and Chang, 2001; Tsochantaridis et al., 2004; Osuna et al., 1997).

2. Proposed algorithm

In this Paper, we consider the problem of segmentation of Leukocyte into Nucleus and Cytoplasm. We classify the Leukocyte into healthy or unhealthy based on various features which are to be extracted from the cell as well as nucleus. So segmentation plays an important role in evaluating the performance of the system. To improve the accuracy of the cell and nucleus segmentation process, we devised a new algorithm which is explained below.

2.1. Segmentation

We need to segment each Leukocyte into Nucleus and Cytoplasm. Segmentation of Nucleus from the Leukocyte is a challenging task and all aspects should be considered clearly. In this research we proposed a new method for the segmentation of Nucleus. The Algorithm for the segmentation is stated below:

The Segmentation Algorithm

Input: X - Input Image

Output: N -Nucleus

```

1  Mat  $a, b, l, h, r1, r2, r3, minf, thresh, morph$ ;
2  string  $s = "X.jpg"$ ;
3   $a = imread(s)$ ;
4   $cvtColor(a, b, CV_BGR2GRAY)$ ;
5   $b.convertTo(l, -1, alpha, beta)$ ;
6   $equalizeHist(b, h)$ ;
7   $r1 = l + h$ ;  $r2 = r1 - h$ ;  $r3 = r1 + r2$ ;
8  Mat  $kernel$ ;
9   $erode(r3, minf, kernel, Point(-1, -1), 3)$ ;
10  $threshold(minf, thresh, 0, 255,$ 
     $THRESH_BINARY_INV + THRESH_OTSU)$ ;
11  $morphologyEx(thresh, morph, 2, el)$ ;
12  $vector<vector<Point>> contours$ ;
13  $vector<Vec4i> hierarchy$ ;
14 Mat  $nucleus = Mat::zeros(morph.size(), CV_8UC3)$ ;
15 Mat  $y = morph.clone()$ ;
16  $findContours(y, contours, hierarchy, CV_RETR_TREE,$ 
     $CV_CHAIN_APPROX_SIMPLE, Point(0, 0))$ ;
17  $int k = maxarea(contours)$ ;
18  $drawContours(nucleus, contours, k, color, -100, 8,$ 
     $hierarchy, 0, Point())$ ;
19  $threshold(nucleus, nucleus, 0, 255, CV_THRESH_BINA$ 
     $RY)$ ;
20  $display nucleus$ ;

```

Initially the color blood slide image 'a' is given as input to the system. The color image is then converted into grayscale image 'l'. To adjust image intensity level, linear contrast stretching is applied to gray scale image l. Then contrast of the grayscale image is enhanced by using histogram equalization method to get image 'h'.

Obtain the image ' $r1=l+h$ ' to brighten all other image components except cell nucleus. Obtain the image ' $r2=r1-h$ ' to highlight the entire image objects along with cell nucleus. Obtain the image ' $r3=r1+r2$ ' to remove all other components of blood with minimum effect of distortion over nucleus. To reduce noise, preserve edges and increase the darkness of the nuclei implement 3-by-3 minimum filter on the image 'r3' to get image 'minf'. Apply a global threshold Otsu's method on image 'minf'.

Using the threshold value in above step convert 'minf' to binary image 'thresh'. To remove small pixel groups use morphological opening and get image 'morph'. We found contours associated with the image. All the contours except the maximum area contour are cleared. The maximum area contour is filled to get image nucleus. The following are the intermediate images that are obtained from the above algorithm.

Fig. 1(a) represents the initial color image. Then Fig. 1(a) is converted to Binary image Fig. 1(b). Then Fig. 1(b) is converted to grayscale image Fig. 1(c). Fig. 1(c) is enhanced by histogram equalization to form Fig. 2(a) and Fig. 2(b) is obtained by adding Fig. 1(c) and Fig. 2(a) and Fig. 2(c) is obtained by subtracting Fig. 2(a) from Fig. 2(b). Fig. 3(a) is obtained by adding Fig. 2(b) and Fig. 2(c). Fig. 3(b) is

the Otsu's threshold of Fig. 3(a) and Fig. 3(c) is the binary image of Fig. 3(b).

Fig. 4(a) is obtained by morphological opening of Fig. 3(c) and Fig. 4(b) is the maximum contour obtained from Fig. 4(a) and Fig. 4(b) is completely filled to obtain Fig. 4(c) which is the final nucleus obtained.

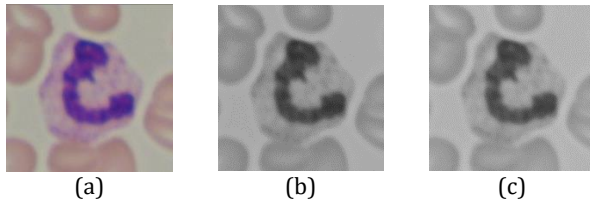


Fig. 1: (a) Color image (b) Binary image (c) Grayscale image

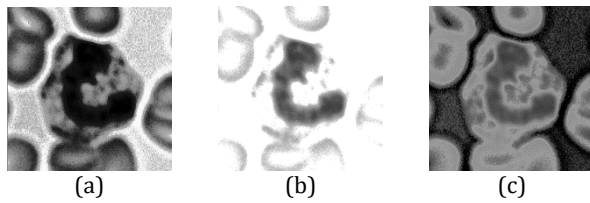


Fig. 2: (a) Histogram image (b) $r1=l+h$ (c) $r2=r1-h$

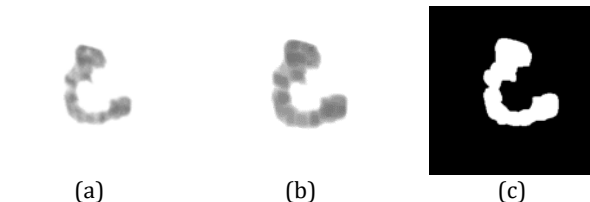


Fig. 3: (a) $r3=r1+r2$ (b) Otsu threshold (c) Binary image

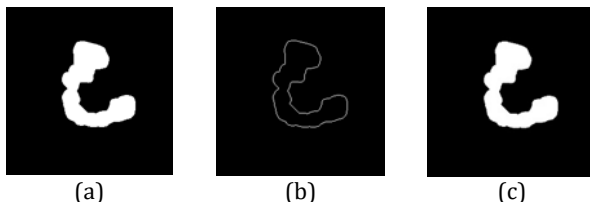


Fig. 4: (a) Morphological opening (b) maximum contour (c) Nucleus

The Segmentation Algorithm (cell)

Input: X- Input Image

Output: C-cell

```

1 Mat a,hsv,tot,x,morph;
2 string s = "X.jpg";
3 a = imread(s);
4 Scalar hsv_l(130, 0, 0),hsv_h(155, 255, 255);
5 cvtColor(a, hsv, CV_BGR2HSV);
6 inRange(hsv, hsv_l, hsv_h, tot);
7 x = tot.clone();
8 threshold(x, x, 0, 255, THRESH_BINARY);
9 x = x + morph;
10 morphologyEx(x, morph, 2, el1);
11 vector<vector<Point>> contours;
12 vector<Vec4i> hierarchy;
13 Mat cell = Mat::zeros(x.size(), CV_8UC3);
14 Mat y = morph.clone();
15 findContours(y, contours, hierarchy,
    CV_RETR_TREE, CV_CHAIN_APPROX_SIMPLE,
    Point(0, 0));

```

```

16 int k=maxarea(contours);
17 drawContours(cell, contours, k, color, -100, 8,
    hierarchy, 0, Point());
18 threshold(cell,cell,0,255,CV_THRESH_BINARY);
19 display cell

```

First color image X is converted to HSV image hsv. Extraction of white blood cell part i.e. image tot is done using HSV values. Convert obtained image to binary image x. To remove small pixel groups use morphological opening. We found contours associated with the image. All the contours except the maximum area contour are cleared. The maximum area contour is filled to get image nucleus. The following are the intermediate images that are obtained from the above algorithm.

Fig. 5(a) is the initial color image taken and then it is converted to Fig. 5(b) which is the HSV image of Fig. 5(a). Then Fig. 5(c) is converted to its binary image Fig. 5(b). Fig. 6(a) is obtained by morphological opening of Fig. 5(c). Fig. 6(b) is the maximum contour obtained and then it is filled to form Fig. 6(c) which is the final cell obtained.

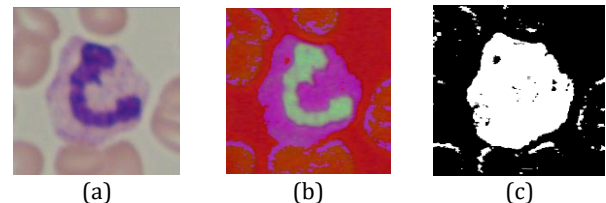


Fig. 5: (a) Color image (b) HSV image (c) Binary image

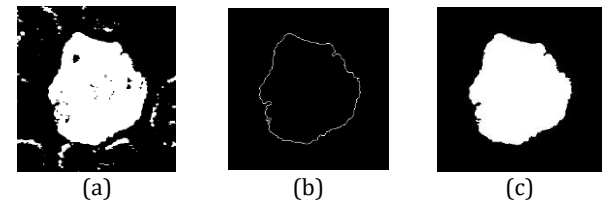


Fig. 6: (a) Morphological opening (b) Maximum contour (c) cell

2.2. Feature extraction

Healthy and Unhealthy Leukocytes can be classified based on certain features. Shape of the Nucleus plays an important role in differentiating unhealthy cells. Once the segmentation process is completed, we extract features like Area, Perimeter, Circularity, Convex area, Solidity, Major axis length, Minor axis length, Eccentricity, Extent, Filled area, Aspect ratio, Equivalent diameter, Mean, Standard deviation, Entropy, Nucleus to cell area ratio, Nucleus to cell perimeter ratio. Definitions of some important features are shown below:

- a) Area: The number of non-zero pixels within the image region is called as area.
- b) Perimeter: The distance between successive boundary pixels is termed as perimeter.
- c) Circularity: This is defined as ratio of area to square of perimeter and then multiplied by 4 times the value of pie. It is dimensionless.

$$\text{Circularity} = 4 * \pi * \text{Area} / \text{Perimeter}^2$$

d) Eccentricity: It is a parameter associated with every conic section. It can be thought of as a measure of how much the conic section deviates from being circular.

e) Form factor: a mathematical factor which compensates for irregularity in the shape of an object, usually the ratio between its volume and that of a regular object of the same breadth and height.

2.3. Classification

After extracting all the required features, we need to classify the Leukocytes into healthy or unhealthy. Classification process plays an important role in detection of leukemia and it's the soul of the entire system. And the accuracy of the system completely depends on the classification algorithm used. Several

machine learning algorithms have been proposed till date for the classification process. Out of them, Support Vector Machine (SVM) is found to be simple and efficient with an accuracy of 92.7%. "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then Classification is performed by finding the hyper-plane that differentiates the two classes very well. After applying SVM on the images, all the Healthy cells fall onto one side of the plane and the Unhealthy cells fall onto other plane. The whole algorithm is illustrated in Fig. 7.

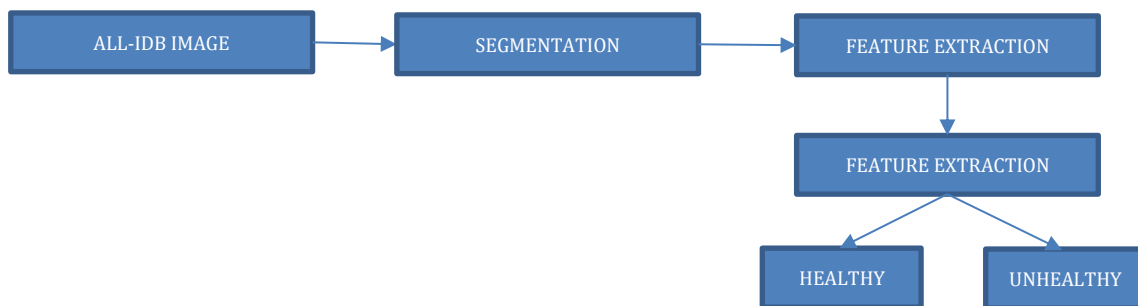


Fig. 7: Flow chart of proposed work

3. Experiment and result

After segmentation of Leukocyte, the results of Healthy cell and Unhealthy cell are shown below. In our experiment 40 images of Healthy cells and 40 images of Unhealthy cells are taken from ALL-IDB2. According to our work, the colored images converted to binary images of both Nucleus and cell (Fig. 8 and 9).

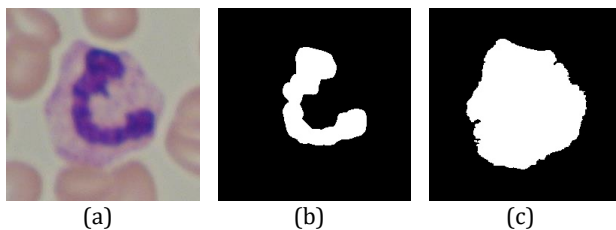


Fig. 8: (a) Healthy cell (b) Nucleus (c) Cell

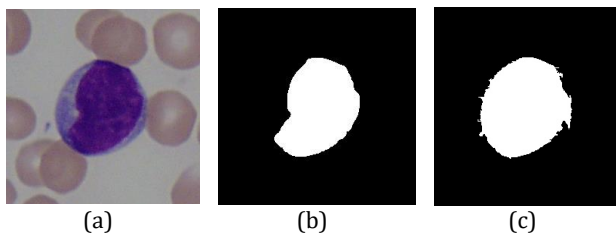


Fig. 9: (a) Unhealthy cell (b) Nucleus (c) Cell

The above images are taken from the experiment results. Both nucleus and cell part are obtained from

the segmentation algorithm (nucleus) and segmentation algorithm (cell) respectively.

Training data are the input feature values which are supplied to the system and experiment is evaluated using the testing data. When training data is given to any classifier it establishes the relation based on the data and then it applies the relation to the test data. In this way, the classifiers work. Overall we took 80 images out of which 40 are healthy and 40 are unhealthy. We compare the results of our work with Khashman and Abbas (2013) method because their method is proved to be the best method available in this field of Leukemia Detection. In the Tables 1 and 2 we present the comparison results and overall efficiency of our work.

Table 1: Comparison with Khashman and Abbas (2013) method

Training and Testing Data Split	Khashman and Abbas	Proposed Method
Training(25%)(10(n):10(a))		
Testing(75%)(30(n):30(a))	75.1	93.33
Training(50%)(20(n):20(a))		
Testing(50%)(20(n):20(a))	80	80

Table 2: Comparison of overall efficiency

Method	Khashman and Abbas	Proposed Method
Overall efficiency	77.55%	86.66%

In Table 1, in the first case training data (25%): test data (75%), 10 healthy images and 10 unhealthy images are given as training data and 30 healthy images and 30 unhealthy images are given for

validation. Then our method achieved accuracy of 93.33% leading the accuracy of Khashman and Abbas (2013) method by 18.23%. In the second case training data (50%): test data (50%), 20 healthy images and 20 unhealthy images are given as training data and then system is evaluated on another 20 healthy and 20 unhealthy images. Then we achieved accuracy of 80% which is equal to the accuracy obtained by Khashman and Abbas (2013) method.

In Table 2, we compared the overall efficiency of our proposed work with the Khashman and Abbas (2013) method and we achieved the accuracy of 86.66% leading their method by 9.11%.

In Khashman and Abbas (2013) method, they used Multi-Layer Perceptron (MLP) method for classification and in our proposed work, we have used SVM method for classification and we proposed a new segmentation process for Nucleus separation. This clearly reveals the strength of proposed segmentation method which provides more efficient nucleus separation to achieve better classification accuracy compared with the existing algorithm.

4. Conclusion

In this paper, we proposed a step by step approach for the detection of Leukemia using Image Processing. The proposed Method provides the best accuracy of 93.33% when 80 images from ALL-IDB2 database are provided as input. Segmentation method used to separate Nucleus from the Leukocyte is simple and effective when compared to all the existing methods. Coming to our future work, In addition to extraction of features from Nucleus, extracting features from Cytoplasm will also help in improving performance of the system. Classification plays an important role in the detection process. So, we intend to implement more Machine learning techniques to improve the accuracy of cell classification algorithm.

References

- Bezdek JC, Ehrlich R, and Full W (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2-3): 191-203.
- Fatma M and Sharma J (2014). Leukemia image segmentation using K-means clustering and HSI color image segmentation. *International Journal of Computer Applications*, 94(12): 6-9.
- Khashman A and Abbas HH (2013). Acute lymphoblastic leukemia identification using blood smear images and a neural classifier. In the 12th International Conference on Artificial Neural Networks: Advances in Computational Intelligence (IWANN '13), Puerto de la Cruz, Tenerife, Spain, 2: 80-87. https://doi.org/10.1007/978-3-642-38682-4_10
- Kuo BC and Landgrebe DA (2004). Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5): 1096-1105.
- Li CH, Kuo BC, and Lin CT (2011). LDA-based clustering algorithm and its application to an unsupervised feature extraction. *IEEE Transactions on Fuzzy Systems*, 19(1): 152-163.
- Majno G and Joris I (2004). *Cells, tissues, and disease: principles of general pathology*. Oxford University Press, Oxford, UK.
- Mohapatra S and Patra D (2010). Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images. In the International Conference on Systems in Medicine and Biology, IEEE, Kharagpur, India. <https://doi.org/10.1109/ICSMB.2010.5735344>
- Mohapatra S, Patra D, and Satpathy S (2012a). Unsupervised blood microscopic image segmentation and leukemia detection using color based clustering. *International Journal of Computer Information Systems and Industrial Management Applications*, 4: 477-485.
- Mohapatra S, Patra D, and Satpathy S (2014). An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. *Neural Computing and Applications*, 24(7-8): 1887-1904.
- Mohapatra S, Patra D, Kumar S, and Satpathy S (2012b). Kernel induced rough c-means clustering for lymphocyte image segmentation. In the 4th International Conference on Intelligent Human Computer Interaction (IHCI), Kharagpur, India, IEEE: 1-6. <https://doi.org/10.1109/IHCI.2012.6481865>
- Osuna E, Freund R, and Girosi F (1997). An improved training algorithm for support vector machines. In the IEEE Workshop Neural Networks for Signal Processing VII, IEEE, Florida, USA: 276-285. <https://doi.org/10.1109/NNSP.1997.622408>
- Patel N and Mishra A (2015). Automated Leukaemia Detection Using Microscopic Images. *Procedia Computer Science*, 58: 635-642.
- Piuri V and Scotti F (2004). Morphological classification of blood leucocytes by microscope images. In the International Conference on Computational Intelligence for Measurement Systems and Applications, IEEE, Boston, USA. <https://doi.org/10.1109/CIMSA.2004.1397242>
- Pui CH and Evans WE (1998). Acute lymphoblastic leukemia. *New England Journal of Medicine*, 339(9): 605-615.
- Shirvoikar M and Virani HG (2016). Detection and segmentation of WBC cells using image processing technique. *International Journal of Technology and Science*, 9(1): 56-58.
- Srisukkharn W, Lepcha P, Hossain MA, Zhang L, Jiang R, and Lim HN (2013). A mobile enabled intelligent scheme to identify blood cancer for remote areas-cell membrane segmentation using marker controlled watershed segmentation phase. In the 7th International Conference on Software, Knowledge, Intelligent Management and Applications, Chiang Mai, Thailand: 104-114.
- Tong S and Chang E (2001). Support vector machine active learning for image retrieval. In the 9th ACM International Conference on Multimedia, ACM, Ottawa, Canada: 107-118. <https://doi.org/10.1145/500141.500159>
- Tsochantaridis I, Hofmann T, Joachims T, and Altun Y (2004). Support vector machine learning for interdependent and structured output spaces. In the 21st International Conference on Machine learning, ACM, Alberta, Canada: 1-8. <https://doi.org/10.1145/1015330.1015341>
- Turgeon ML (2005). *Clinical hematology: Theory and procedures*. Lippincott Williams and Wilkins, Pennsylvania, USA.
- Vaghela HP, Modi H, Pandya M, and Potdar MB (2015). Leukemia detection using digital image processing techniques. *Leukemia*, 10(1): 43-51.
- Wu KL, Yu J, and Yang MS (2005). A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests. *Pattern Recognition Letters*, 26(5): 639-652.
- Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, and Cheng C (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, 1(2): 133-143.