ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

Analysis of Blood Smear Images for Leukaemia Detection Using Data Mining Algorithms

[1] Vasundhara Acharya, [2] Dr.Preetham Kumar
[1] Student, [2]Professor and Head,
[1][2]Dept. Information and Communication Technology, MIT,
Manipal, India
[1] maithreyi3@gmail.com [2] preetham.kumar@manipal.edu

Abstract: **-- Blood cell segmentation and extraction of characteristics of cells plays a vital role in the field of medicine. Blood count is used to analyze the overall health of a person. It can be used to identify diseases like anemia, infection and leukemia. White Blood cells, Red Blood cells and platelets forms the parts of blood. In laboratory, blood cell counting is done by using counting chamber known as Hemocytometer, Petri dish and microscope. The entire procedure involves the use of physician's skills to prepare the plating which is very time consuming and inaccurate task. The aim of this research is to perform a survey on computer aided system that can detect and estimate the number of red blood cells and white blood cells in the blood smear image using image processing algorithms. Image processing algorithms involves the following steps: Inputting the image, pre-processing, image segmentation, feature extraction and applying appropriate counting algorithm. The main objective here is to gain knowledge about the different methodologies used for counting of red blood cells and extracting features of white blood cells. It also throw light on various research directions used.**

IndexTerms: **—Haemocytometer,Hough-Transform,K-MeansClustering,Petridish,Regionprops**

I. INTRODUCTION

Leukemia is cancer of the blood or bone marrow. It leads to production of large number of white blood cells which are abnormal. The different types of leukemia are: Acute Lymphoblastic Leukemia, Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic myeloid leukemia (CML) [16].The observation of blood samples by Lab physicians is one of the main procedures available for the identification of different diseases like leukemia and anaemia.The humans capability of visualizing the cell features is time consuming and is less precise. The results will be highly dependent upon the skills of the operator. These reasons have led to the use of computer systems to perform the same work. The advantage of using computer system for analysis requires only one image and not the entire blood sample. The precision of result obtained using computer system is much higher compared to the result obtained by human analysis.

In the market there are numerous systems for the automatic calculation of blood cells present in the blood smear. These counters make use of techniques of flow cytometer to measure physical characteristics and chemical properties of the blood cells [17]. The precision of calculation of number of cells may be high but it fails to detect the abnormalities in the cell such as eccentricity which plays a vital role in differentiating a cancerous cell from a non-cancerous one. The use of image processing techniques have been growing these years. The techniques vary depending upon the type of the cell to be detected. A typical blood image usually has the following components: Red Blood cells, White Blood cells and Platelets. This research paper is intended to perform a study on researches conducted for achieving the goal of detection, feature extraction and counting of Red Blood cells and White Blood cells. It also throw light on different feature extraction techniques.

II. LITERATURE SURVEY

a. Techniques for Detection of White Blood

1.1 WBC Segmentation Based On Multi scale Analysis

Belekar [1] in their study have used the mathematical morphology and SMMT operator for segmentation of nucleus and cytoplasm. For the segmentation of nucleus they have utilized two approaches that are: level set method and Watershed

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

algorithm. After the application of two approaches they have compared the results and identified that application of SMMT operator would improve the results. They have made use of SMMT operator along with threshold operation to prevent the problem of leakage. Erosion technique was used to remove small particles. For the detection of cytoplasm they used granulometric analysis and mathematical morphology technique. A comparison between the two approaches was also done. Watershed transform was used to separate the touching objects. Level set method represents surface as a level zero interface of a higher dimensional function called Level-Set function. It could easily deal with the changes in topology and discontinuities in the evolution of the level zero curves. The paper was able to detect the nucleus and cytoplasm but no details about the quantification of cells or extraction of features was specified.
.

1.2 Segmentation by applying pre-processing at various stages

Lorenzo [2] in their work have avoided the different steps of pre-processing and segmentation. They have applied pre-processing at each stage of segmentation that provided a simple and robust solution for segmentation of WBC from other cells. At the first step the leukocytes were identified. The background was identified using threshold value. They used a threshold value based on triangle method or Zack algorithm. The triangle method took the image histogram as the input. It then constructed a straight line between highest peak and lowest peak. The distance between the line drawn and the peaks was calculated. The threshold value was the point at which the distance “d” had the maximum value. The background image obtained was then subtracted from the original image. The unwanted small particles were removed by performing area opening with a structuring element. The leukocytes that were grouped were removed by considering the roundness values of the connected components. If the roundness value was greater than a certain predetermined threshold they were individual leukocytes and did not require any separation. But if it was less than prescribed threshold then it requires the separation .Separation was performed by getting the information of the points of concavity and the information related to the points of maximum image in gray tones. A cutting line that fits the contour was drawn. For the detection of cytoplasm the image was cropped to obtain individual leukocytes. The above procedure could be used only when there is valley of high and low peak created. The process of cropping was a time consuming work.

1.3 Segmentation using Otsu’s threshold algorithm

Minal [3] in their study have proposed an algorithm for segmentation of white blood cells using global Otsu's threshold. The database was taken from ALL-IDB with a size of 108 images. The image after the acquisition was converted to gray scale image. Histogram equalization method was applied to it. After the equalization subtraction operations were applied on the images. The unwanted objects were eliminated using morphological operations. A 3X3 median filter was applied on the resulting image. Global Otsu's threshold was applied on it. The resulting image was converted in to binary and the smaller particles were eliminated using opening operation. A size test was performed and the objects with area which is less than 50 percent was eliminated. The resulting image was used for the purpose of feature extraction. Three features of WBC were observed. They were Area, Perimeter and circularity. The classification of cells was done by using the K nearest neighbor algorithm. The paper presents a new approach for detection of White blood cells, but it has not specified any approach to separate the touching cells. The touching cells may provide a wrong result while quantization of the cells. Only three features were taken in to consideration for classification of lymphoblast. The feature like eccentricity which is an indicator of lymphoblast was not calculated.

1.4 Segmentation using Binarization and local maxima generation

Meng [4] in their work proposed a novel cell detection methodology that takes in to consideration both intensity and shape information of cell to improve the quality of segmentation. The data used in this paper were HeLa cell line. Pictures of the cell nuclei were clicked every 15 minutes with an automated time-lapse fluorescence microscopy. The dataset had 240 images. An online algorithm for classification was also proposed. The online algorithm was adaptive to the changing experimental conditions. An Online Support Vector Classifier (OSVC) was put forward which would contribute to the removal of support vectors from the old model and would assign the new training examples with different weights according to their varying importance.

IFERP
connecting engineers... developing research

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

A binarization process was done to obtain the cell shape information. Next, both intensity and shape information were used for the process of local maxima generation. At last the pixels of nuclei were made to move in to the gradient vector field which would later converge at those local maxima. For the separation of the touching objects a seeded watershed algorithm was applied to it. Prior to the binarization, the cell clusters were separated from the background. A data driven background correcting algorithm which used the cubic B-spline was used to detect the background. In local maxima generation, the original image was added with the image obtained after the distance transformation. A Gaussian filter was applied to it. Techniques for noise removal were applied. The local maxima was generated to that image. The local maxima obtained were the cell centers. The cell detection problem was reduced in to detecting local maxima in the gradient image. The threshold was calculated by using the number of pixels in the original cell. In each local region, only one local maximum was treated as the cell center. The cells were then segmented using the seeded watershed algorithm using the local maxima as the seeds. The result showed that seeded watershed algorithm had less over segmentation problem as compared to others. A comparison between the SVM (Support Vector machine), LASVM, OSVC was done. The result showed that OSVC had an upper hand over the other two.

1.5 Segmentation using Threshold Technique

Farnoosh [5] in their study have proposed a framework for detection of WBCs.The data set used had 20 images. The color scale image obtained was converted to gray scale. The images were cropped to have one WBC per image. Edge detection was performed to detect the edges of nucleus .They identified that Sharp changes in image brightness played a very important role in boundary detection. Points in the image where brightness changes significantly were often referred to as edges or edge points. Next, GVF deformable contour was done with suitable iterations. Snake algorithm found the connected boundary that was detected and it selected the nucleus. Fill operation was performed to fill holes that could not be reached from the boundary. In order to detect the cytoplasm the nucleus detected was subtracted from the original image. Next the cytoplasm was segmented using the Zack algorithm. In this paper they have detected the nucleus and cytoplasm but it involves the cropping of images to sub images which is a time consuming task.

Yazan [6] in their study proposed an approach for segmentation of RBCs and WBCs using threshold technique. The image was converted to gray scale to eliminate its Hue and Saturation and retain luminance.Next,it was converted to binary by using a threshold value of 64.The resulting image had WBCs in black and background was white. So it was complemented. The unwanted small particles were removed using morphological open operator. This image was eroded to reduce the number of overlapping cells. A fill operation was done to fill holes. They identified that fill operation could improve the results of edge detection. In this paper, a fixed threshold of value 64 is used. It satisfies only some images and produces inaccurate results.

1.6 Segmentation Using Fuzzy Logic

Bhagavathi [7] in their work have used the Fuzzy logic for the segmentation of WBC.Initially, the RGB image obtained was converted using gray scale image. They identified that RBC's were rich in Red color. They used a 'adapthiseqfuction' to highlight the red components. Next the edges were detected using the fuzzy rules. Rules were defined for the fuzzy inference system.FIS editor was created with input values and rules were set to detect the circles. In order to detect the concentric circles multiple radiant arguments was used. The result was an accumulator array. The resulting image had white background and black boundary. In order to achieve better identification the image was complemented. The holes which could not be reached from the object boundary were filled and labeling was performed. Based on the intensity values the WBCs were separated from RBCs.Erosion was applied and WBC mask was obtained. In this paper, while detecting WBCs it does not accurately determine the WBC mask for some images which produces inaccurate results while calculating the feature like area.

1.7 Segmentation Using K-means Clustering

Vinod [8] in their study proposed an approach for identification of WBCs by using K-means clustering.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

The images of bone marrow were obtained from Internet. Next, in the preprocessing step the noise was removed using the threshold technique. For the segmentation purpose, they have utilized k-means clustering algorithm. Number of clusters chosen was 4.In one cluster it would hold the WBC, another would hold the RBC, next would hold the cytoplasm and last would hold the background. Based on WBC extracted, a number of features such as area, perimeter, eccentricity and some texture features like energy, homogeneity were calculated. For the purpose of calculation they used Support Vector Machine. In this paper, they have not specified the details about the results obtained. Even though they have used four clusters it does not separate cytoplasm from the RBC for all images. Hence a second step of processing to clearly separate RBCs from the cytoplasm is required.

Seeedizadeh [9] in their work proposed an approach for detection of WBCs using the bottle neck algorithm and K-means. The dataset contained 50 images of bone marrow collected from Alzahra hospital. The input image was converted from RGB to HSV color space,. The resulting image was enhanced by using contrast stretching. A median filter was applied to remove salt and pepper noise from it. After this step, in order to detect cells, a segmentation algorithm was applied which worked based on difference in color of RBC, WBC and background. Next, the mask of WBC was obtained by using the Otsu's threshold algorithm. A morphological closing operation with a 3 pixel diameter and morphological open operation with a 4 pixel diameter was applied on it. Objects with area less than 500 pixels were eliminated. The resulting object was the cell mask. To obtain the nucleus mask the image was converted to LAB color space. The same procedure was applied, but in the open operation the objects with an area less than 100 pixels were eliminated. The mask obtained formed the nucleus mask. The clumped nuclei were eliminated using the bottle neck algorithm. A pair of points was chosen for splitting. Next step was to join the split points. They then calculated the Area and Convex area. They then calculated the ratio between Area and Convex Area. They identified that if area was greater than 1000 pixels and if the ratio was less than 0.9 then it was a touching nuclei and it required the application of bottle neck algorithm. Features like nucleus to cytoplasm ratio and eccentricity was used to classify the normal cell from other marrow cells. They calculated the means of a and b components of LAB color space. If it would go out of certain range, they rejected it.

b. Techniques to Detect and Count Red Blood cell

2.1 Segmentation and Identification of Red blood cells using Threshold Technique

Pooja [10] in their work proposed an approach for identification and counting of RBCs. They identified the disadvantages involved in the counting done manually. The dataset had images collected from 15 patients. The input image was converted to gray scale. A median filter was applied to the resulting image was applied .Then Otsu's threshold was applied on the image. The fill operation was performed in order to segment the blood cells properly. The cells at the border were cleared as those would not contribute much to the calculation. The resulting image was labeled with two specific values which were 4 and 8.They obtained an accuracy of 94.58 percent. They also provided details about the approach followed in the manual count of the RBCs. In this paper, Otsu’s threshold is used which may produce wrong segmentation if the threshold determined is incorrect.

2.2 Segmentation and Identification of Red blood cells using Fuzzy Logic

Bhagavathi [7] in their work proposed an approach for detection and counting of RBCs.The input image was applied with ‘adapthiseqfuction’ to highlight the red components. The resulting image was used for edge detection. For the purpose of edge detection the fuzzy rules were framed. An FIS Editor was created based on the input values. Rules were set to detect the circles. They assumed that if the radius of the circle fell within the range ,then it was identified as pixel 1.If in case the radius falls out of the range then it was identified as pixel 0.For the detection of the concentric circles multiple radiant argument was used. The output of the above operations was an accumulator array with same dimensions as the input image. The resulting image had white background and black boundary. The image was then complemented. Hole filling operation was performed. Based on the intensities WBCs were separated from the RBCs.Circle drawing algorithm was used. The results section of the paper shows that while applying the circle drawing algorithm, it also considers the WBCs as RBCs.In order to avoid such errors a second stage of processing is required to separate the WBCs from RBCs.

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 3, Issue 11, November 2016

2.3 Segmentation and Identification of Red blood cells using Hough transform method.

Mausumi [11] in their study proposed a methodology for counting of RBC using Hough transform method. The dataset had 5 images. The image was made to undergo a number of steps. The image was converted to gray scale image first. The 'adpathiseqfunction' was applied to it. The resulting image was given for edge detection. Spatial smoothing filtering was applied. Hough transform method was used to separate the RBCs from WBCs based on their sizes and shapes. After the separation of RBCs from WBCs a counter was applied to it. A formula was derived to calculate the number of cells. A comparison of number of cells counted automatically and manually was done. In this paper, while applying the circle detection algorithm it detected even the WBCs as RBCs.Hence a second processing step to separate the RBCs from WBCs is required.

Thejashwini [12] in their work proposed a technique to identify and count RBCs.The first step was to separate the red and blue component and consider only green component of the image. Next step was to apply the contrast adjustment. The resulting image was given as an input to the circular Hough transform technique. It would identify and detect the cells. A counting operation was done to count the number of circles drawn which would depict the number of RBCs in the smear. In this paper, the approach used by them is applicable only if shape of WBCs is not circular. If it is circular then it misinterprets the WBCs as RBCs.Another step to separate the WBCs from RBCs is required.

Siti [13] in their work proposed an approach to count the RBC using Hough transform. The input image was cropped in to sub images to obtain the minimum and maximum radius of RBCs.The sub image was converted to gray scale. Morphological operations like opening, closing and reconstruction were applied to it. The resulting image was converted to binary. The border of the image was cleared. The measures like minimum radius, maximum radius, standard deviation were calculated. Circular Hough transform was used with the radius calculated from previous step as input.

Venkatalakshmi [14] presented a method for automatic red blood cell counting using Hough transform. The input image was converted to HSV format. The S component of the image was extracted for the further analysis. Lower and Upper threshold were calculated from the histogram. Based on the threshold values the image was divided in to two. An area closing was applied to lower pixel value image and morphological dilation and area closing was applied to higher pixel value image. OR operation was performed on the resulting images and Circular Hough transform was applied to extract out the RBCs.

2.4 Segmentation and Identification of Red blood cells using Watershed Transform

Hemant [15] in their work proposed a method to extract RBCs and count them using Watershed Transform. The combination of Ycbcr with morphological operations were used to achieve it.In the preprocessing step the image was freed from the noise and smoothing was applied. The image was converted to Ycbcr format. Second component of the result was obtained. They identified that Ycbcr format overcame the problem of illumination. Morphological opening with a specific structuring element was used. The resulting image had both WBCs and platelets. Another step of subtraction and opening was performed to obtain two separate images with WBCs and Platelets. To obtain only RBCs, the input image was converted to gray scale and a open by reconstruction operation was used. Erosion operation was applied to the resulting image. Next a closing by reconstruction was applied to it. The binary image having all the cells is obtained. The image having the WBC mask was subtracted from the image having all the cells. The output image had only RBCs.The overlapping cells were separated from each other using the Watershed transform. A final count operation was done to count the cells.

III. CONCLUSION

In this paper, discussion about the different methods to detect and segment blood cell in to different parts like RBCs and WBC is done. At the beginning the different approaches used to segment the WBCs is discussed. Each of the approach used had its own advantage. The algorithms varied depending upon the use of the morphological operators and threshold technique. In the second section, discussion about the different approaches to segment the RBCs is done. No single approach is enough to completely detect the WBCs and

RBCs.In future, it requires the combination of approaches to completely automate and accurately detect the WBCs and RBCs.

REFERENCES

[1] Belekar S.J and Dr. Chougule .S .R ,A"WBC Segmentation Using Morphological Operation and SMMT Operator-A review",International Journal of Innovative Research in Computer and Communication Engineering , 2015.

[2] Lorenzo Putzu, and Cecilia Di Ruberto "White Blood Cells Identification and Counting from Microscopic Blood Image ", International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering , Vol.7,No. 1, pp.20-27, 2013.

[3] Minal D. Joshi , Atul H. Karode and S.R.Suralkar "White Blood Cells Segmentation and Classification to Detect Acute Leukemia",International Journal of Emerging Trends & Technology in Computer Science, Vol.2,No.3, pp.147-151, 2013.

[4] Meng Wang , Xiaobo Zhou1, Fuhai Li ,Jeremy Huckins ,Randy W King and Stephen T.C. Wong,"Novel Cell Segmentation And Online Learning Algorithmd For Cell Phase Identification in Automated Time-Lapse Microscopy",Institute of Electrical and Electronics Engineers, pp.65-68, 2007.

[5] Farnoosh Sadeghian,Zainina Seman,Abdul Rahman Ramli,Badrul Hisham ,Abdul Kahar and M-Iqbal Saripan ,"A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing",Biological Procedures Online, Vol.11,No.1,pp.196-206,2009.

[6] Yazan M. Alomari ,Siti Norul, Huda Sheikh Abdullah ,Raja Zaharatul Azma and Khairuddin Omar "Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm", Computational and Mathematical Methods in Medicine, pp.1-14, 2014.

[7] Bhagavathi S. L. and Thomas Niba S. "An Automatic System for Dectecting and Counting RBC and WBC Using Fuzzy Logic", ARPN Journal of Engineering and Applied Sciences ,Vol.11,No.11,pp.6891 – 6894, 2016.

[8] Vinod Mulik ,Sonam Alhat ,Poonam Bhilare and Vijayalaxmi Bhoge "Analysis of Acute Lymphoblastic Leukemia Cells using Digital Image Processing", International Journal for Scientific Research and Development, Vol .4,No.2,pp.70-72, 2016.

[9] Z.Saeedizadeh ,A.Mehri Dehnavi ,A.Talebi and H.Rabbani "Automatic recognition of myeloma cells in microscopic images using bottleneck algorithm, modified watershed and SVM classifier", Journal of Microscopy,Vol.261,No.1,pp. 46–56,2015.

[10] Pooja R. Patil ,G.S. Sable and Gauri Anandgaonkar "Counting of WBCs and RBCs from Blood Images Using Gray Thresholding "International Journal of Research in Engineering and Technology,Vol. 3,No.4,pp.391-395,2014.

[11] Mausumi Maitra,Rahul Kumar Gupta and Manali Mukherjee "Detection and Counting of Red Blood Cells in Blood Cell Images using Hough Transform",International Journal of Computer Applications ,Vol.53,No.16,pp.18-22, 2012.

[12] Thejashwini M and M C Padma "Counting of RBC's and WBC's Using Image Processing Technique", International Journal on Recent and Innovation Trends in Computing and Communication,Vol.3,No.5,pp.2948-2953, 2014.

[13] Siti Madihah Mazalan, Nasrul Humaimi Mahmood and Mohd Azhar Abdul Razak "Automated Red Blood Cells Counting in Peripheral Blood Smear Image Using Circular Hough Artificial Intelligence, Modelling & Simulation ,pp.285-289,2013.

[14] Venkatalakshmi. B and Thilagavathi. K, "Automatic Red Blood Cell Counting Using Hough Transform",Proceedings of IEEE Conference on Information and Communication Technologies,pp.268-270,2013.

[15] Hemant Tulsani and Naveen Yadav and Saransh Saxena "Segmentation using Morphological Watershed

Transformation for Counting Blood Cells",International Journal of Computer Applications and Information Technology,Vol.2,No.3,pp.28-36,2013.

[16]"Hemotology":http://www.hematologyatlas.com/principalpage.htm

[17]"Leukaemia".http://patient.info/in/health/leukaemia-a-general-overview.