International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 6 (March 2019) - Special Issue www.ijirae.com
SPECIAL ISSUE - 5th Inter National Level Conference - "MEEMIC – 2019"

IOT BASED AUTOMATED DETECTION OF WBC CANCER DISEASES

Premavathi M, Rajbrindha M, Srimathi P, Swetha A
Guided by
Dr.R.Satish Kumar , Principal
Sengunthar College of Engineering, Tiruchengode.
principal@scew.org, premamuthusamy07, rajbrindha97, srimathiprakasam, ananthswetha2203 {@gmail.com}

Manuscript History
Number: **IJIRAE/RS/Vol.06/Issue03/Special Issue/SI.MRAE10110**
Received: 20, February 2019
Final Correction: 05, March 2019
Final Accepted: 20, March 2019
Published: **March 2019**
Editor: Dr.A.Arul L.S, Chief Editor, IJIRAE, AM Publications, India

Abstract: Acute Myeloid Leukemia (AML) is one of cancer type that attack White Blood Cells in myeloid descendants. On the clinical examination of leukemia, the number of each blast cell in the laboratory is calculated. However, in some subtype of AML like M4, M5 and M7 are affected by the same type of precursor cells. The precursor cell of them is myeloblast, monoblast and mega karyoblast, which needs more detailed analysis to distinguish. Classification is performed on cell types of precursors cells derived from bone marrow preparations. The stages that have been completed are preprocessing, segmentation, extraction and feature selection, and classification. Features used as input of classification stage are area, nucleus ratio, circularity, perimeter, mean, and standard deviation. The support vector machine classification results in the best performance test data are achieved by linear kernel. The performance was obtained by combining six features for eight cell types from the maturation of the three precursor cells. This diagnosed detail will be updated on cloud by IOT module.

Keywords—acute myeloid leukemia, bone marrow, segmentation, classification, support vector machine

I. INTRODUCTION

Leukemia is one of the blood cancers that attack white blood cells that form in the bone marrow. In patients with leukemia, the bone marrow produces white blood cells (blast) excessively. Excessive white blood cells will cause the accumulation of young white blood cells in the bone marrow. It can affect the inhibition and decrease of healthy blood cells. The French-American-British (FAB) hematologic classification system divides leukemia into four types based on its forming cells; they are Acute Myeloid Leukemia (AML), Acute Lymphocytic Leukemia (ALL), Chronic Myeloid Leukemia (CML) and Chronic Lymphocytic Leukemia (CLL) . Acute leukemia is characterized by the rapid development of blast cells in the blood. If not treated immediately, it can lead to death in a matter of weeks or even days. Furthermore, acute leukemia is divided into eight subtypes; they are M0, M1, M2, M3, M4, M5, M6, M7. In some subtype of AML like M4, M5 and M7 are affected by the same type of precursor cells.

The precursor cell of them are myeloblast, monoblast and megakaryoblast , which needs more detailed analysis to distinguish. Generally, leukemia is identified by counting the number of white blood cells and red blood cells through a microscope based on cell morphology by hematologists. However, the procedure of calculating blood cells under a microscope is still relatively time-consuming, highly dependent on the operator's ability and fatigue factor. This research aims to classify cells based on cell type on AML subtypes M4, M5 and M7. The techniques used are the K-Means and some additional methods of segmentation and also multiclass Support Vector Machine using one-vs-rest comparison model for cell of the development of precursor cells. The results of this study are expected to reduce errors and inconsistencies in classification of cell types in AML subtypes M4, M5 and M7.

II. RELATED RESEARCH

Some studies have done a lot of enhancement technique of image quality in recent years. One of them has been done by using contrast stretching technique. Contrast stretching is used to improve the process of diagnosis of acute leukemia images, thus providing additional information on the cytoplasm and nucleus cell. Research on the improvement of image quality among others has been done by using median filter. Median filter could provide a smoother appearance by retaining the edge detail of the object. At the stage of identification of medical object, one of the important stages is segmentation. Segmentation aims to separate the part of the cell body consisting of nucleus and cytoplasm. The thresholding method also can be used for segmentation. On the other hand, another technique use color as a medium for object separation by using K-Means. Besides these two methods, watershed distance transform also can be used to perform separation of stacked cells. Morphology reconstruction technique is also applied with good result for segmentation. Feature extraction has an important role in the classification technique as the stage performed after the segmentation process. Features are used as input values that represent the circularity are very influential in determining the characteristics of the object. It is proved by a maximum accuracy, value of 95.70%. Other features such as area, ratio of nucleus, recognize objects with an average accuracy of 89.68%. In addition to these four features, color features like mean and standard deviation of RGB colors can also be applied. The classification stage of objects based on features has been done by using supervised classification techniques with support vector machine. Classification with SVM techniques can be applied linear non-linearity and are very powerful for data splitting based on hyperplane classifier. The SVM technique has been used to separate objects by applying linear separation with an accuracy of 93.5% and used to separate six cell types with an accuracy of 97%.

METHODOLOGY

A. Image Acquisition

AML M4, M5 and M7 preparations were obtained from RSUP Sardjito Yogyakarta through Ethical Approval procedure. Images are captured using a 21-megapixel resolution digital camera attached to Olympus ocular lens with 1000 times magnification. Image retrieval was performed 35 times for each subtype.

B. Image Enhancement and Segmentation

Image enhancement was done by applying median filtering and global contrast stretching by considering all minimum and maximum color range for all RGB channels. The combination of colors will provide a minimum and minimum value that will be used as a contrast image illumination process. The segmentation stage consists of two parts, they are; whole cell segmentation and segmentation of the nucleus of each cell. On whole cell segmentation, the enhanced image is separated into two as marker and mask images for reconstructing process. The mask image determination process uses 2 color cluster models in the K-Means process. The first cluster model is with a* and b* colors for AML M4 and M5, while the second cluster model applied to AML M7 with color L and b*. Watershed distance transform (WDT) technique also applied for cell separation. The flowchart of whole cell segmentation is shown in Fig.1.1. Segmentation of the nucleus in each cell is performed to calculate the cell ratio feature. The RGB image of the nucleus must be converted first to the HSV color model because the nucleus object has a high saturation value which contributes to the separation of the object. The nucleus segmentation flowchart is shown in Fig.1.2.

C. Feature Extraction

The feature extraction process is performed to obtain six numerical data that representing cell characteristics, including area, perimeter, circularity, nucleus ratio, mean and standard of deviation. The six features are used as input on the SVM classifier as follow:

- **Area:** Defines the number of white pixels from binary segmentation cells in the image [17], defined by Equation(1).

$$A=\sum_{x=0}^{n-1}\sum_{y=0}^{n-1}\left[i(x,y)i[(x,y)=1,f(x,y)]\right] \quad (1)$$

  Variable *i(x, y)* is the pixel value in the image *f(x, y)*. Variable *A* defines the area. Variable *n* is the number of the length of image row, whereas *m* is the length of image column.
- **Perimeter:** Defines the number of white pixel edges of a binary segmentation cell with one-pixel structuring element, defined by Equation(2).

$$P=A-(A\emptyset B) \quad (2)$$

  Variable *B* is a strel of image morphology. Variable *A* is the area of the object, and the *P* is the perimeter of the object.
- **Circularity:** Defines a circular level of objects in the range of values 0 to 1, defined by Equation(3).

$$C=4nA/P2 \quad (3)$$

  Variable *A* is the area of whole cell, whereas *P* is the perimeter of *A*. The more rounded an object, the closer the value of 1.

- **Nucleus Ratio:** Defines comparison between the extent of nucleus and whole cell in the range of values 0 to 1, defined by Equation(4).

$$R=(A(nucleus))/(A(cell)) \quad (4)$$

Variable *A(nucleus)* defines the area of nucleus object, while *A(cell)* defines the whole area of each cell. *R* is the ratio of the nucleus area.

- **Mean:** Defines the average of color intensity of red, green and blue channels form objects, defined by Equation (5).

$$x' = 1/(n.n)\sum_{x=0}^{n-1}\sum_{y=0}^{n-1} i(x,y) \quad (5)$$

Variable *i(x, y)* is the pixel value. Variable *n* is the number of the length of image row whereas *m* is the length of image column. Variable defines the average of pixel intensity value of the cell.

- **Standard Deviation:** Defines the magnitude of the difference in the sample data value to the mean value, defined by the Equation(6).

$$o = \frac{1}{n.n}\ (\sum_{x=0}^{n-1}\sum_{y=}^{n-1} i(x,y) - x'))\text{^}2 \quad (6)$$

Variable *i(x, y)* is the pixel value. Variable *n* is the number of the length of image row, whereas *m* is the length of image column. Variable defines the average of pixel intensity value and o is the standard deviation of the cell.

D. Normalisation and feature combination

Numerical data from feature extraction has very varied range that needs to be normalized. Normalisation was done to make the feature scaled uniformly in the 0 to 1 value range. Before being processed at the classification stage, the data firstly processed with Equation (7).

$$Normalization = (data[x]-min(data)) /(max(data)-min(data)) \quad (7)$$

Features obtained from normalization process will be a subject to feature selection process with Minimal Overlap Probability (MOP). The process was done to determine the best combination of parameters of the six features, which is calculated by Equation(8).

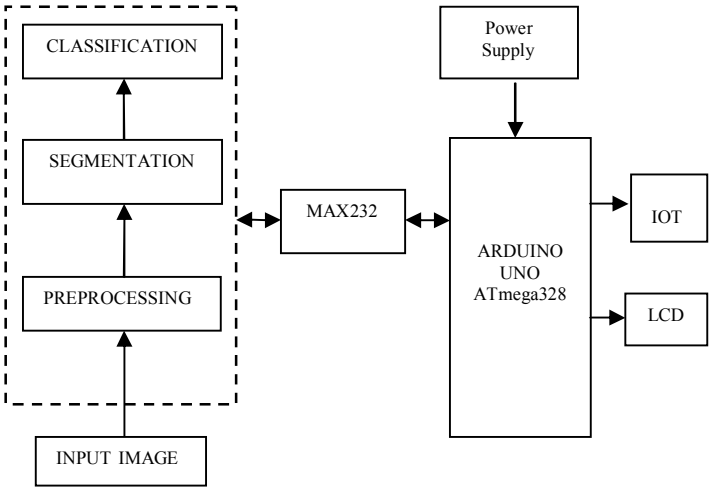
$$ProbError=(2*Lo)/(Ls+Lb) \quad (8)$$

Ls is the first feature range, and Lb is the second feature range. Whereas Lo represents the difference in value between boundary value of Ls which intersects with Lb [18]. Selection of feature combinations is based on the smallest Probability Error (ProbError) value difference.

E. Training and Testing

The classification process begins by applying parameter selection with grid search. The training model used is one-vs- rest, by looking at the class y1 (one subtype) as the positive class and the class x1 (other) as the negative class. The training model was set up k times according to the number of cell types. The training process is carried out to find the best hyper plane based on data features, data labels and parameters which act as input model in testing phase. The percentage of training and testing data applied was 80%: 20%

III. BLOCK DIAGRAM



4.1 ARDUINO UNO ATmega328 : Arduino Uno is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with USB cable or power it with a AC to DC adapter or battery to get started wrong, worst case scenario you can replace the chip for a few dollars and start over again. Uno means one in Italian and was chosen to mark the release of Arduino Software (IDE) 1.0. The UNO board and version 1.0 of Arduino Software (IDE) were the reference versions of Arduino, now evolved to newer releases. The Uno board is the first in a series of USB Arduino boards, and the references model for the Arduino platform, for an extensive list of current, past or outdated boards see the Arduino index boards.

4.2 LCD : A liquid-crystal display (LCD) is a flat panel display, electronic visual display or video display that uses the light modulating properties of liquid crystals. Liquid crystals do not emit light directly. LCDs are available to display arbitrary images (as in a general purpose computer display) or fixed images which can be displayed or hidden, such as preset words, digits, and 7-segment displays as in a digital clock. They use the same basic technology, except the arbitrary images are made up of large number if small pixels, while other displays have larger elements. A liquid crystal display or LCD draws its definition from its name it is combination of two states of matter, the solid and the liquid. LCD screen is more energy efficient and can be disposed of more safely than CRT.

4.3 IOT : The Internet of Things (IoT), also sometimes referred to as the Internet of Everything (IoE), consists of all the web-enabled devices that collect, send and act on data they acquire from their surrounding environments using embedded sensors, processors and communication hardware. These devices, often called "connected" or "smart" devices, can sometimes talk to other related devices, a process called machine-to-machine (M2M) communication, and act on the information they get from one another. Humans can interact with the gadgets to set them up, give them instructions or access the data, but the devices do most of the work on their own without human intervention. Their existence has been made possible by all the tiny mobile components that are available these days, as well as the always-online nature of our home and business networks.

V. RESULT DISCUSSION

This research applied computation and simulation of data features from all segmented cell form all image. Computation and simulation are implemented using Matlab by predefined methodological steps.

A. CELL TYPE IDENTIFICATION

The obtained images of AML M4, M5 and M7 consists of eight cell types of different amounts. These cell types are obtained by expert identification support. The eight cells are derived from 3 precursor cells, they are myeloblast, monoblasts and megakaryoblast. These three undergo a process of maturation with their respective cell. Myeloblast develop into 3 types, namely myeloblast, promyelosit and granulocytes. Monoblast develop into monoblast, promonosit, monocytes, whereas into seven cells, there is a lymphocyte cell derived from lymphoblast stem cells encountered in all identified AML subtypes. A sample of each cell is shown in table 5.1.

No.	Cell Type	Image of Object
1.	Myeloblast	
2.	Promyelocyte	
3.	Granulocyte	
4.	Monoblast	
5.	Promonocyte	
6.	Monocyte	
7.	Megakaryoblast	
8	Support (Lymphocyte)	

Table 5.1 sample of each cell

B. CELL DEGMENTATION RESULTS

The segmentation results show that from the total of 105 images consisting of 35 images of each subtype, 1500 cells are correctly segmented, and 210 cells are incorrectly segmented. The detailed comparison of cell numbers of each subtype is shown in Table 5.2. Moreover, sample of cell segmentation results in each subtype are shown in Table 5.3

Type	Detail			
	Correct Segmentation	Incorrect Segmentation	Total Cell	Percentage
AML M4 Preparation	300	41	341	87.98%
AML M5 Preparation	448	134	582	76.97%
AML M7 Preparation	752	35	787	95.55%
Total	1500	210	1710	87.72%

TABLE 5.2.DETAILS OF SEGMENTATION RESULT OF EACH SUBTYPE

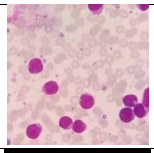
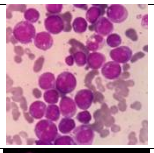
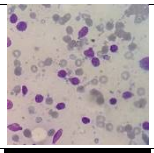
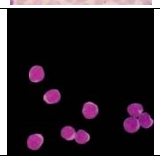
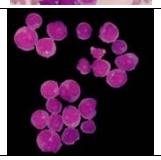
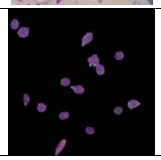
Image Type	Subtype		
	AML M4	AML M5	AML M7
Original Image			
Segmentation Results			

TABLE5.3.SAMPLE OF WHOLE CELL SEGMENTATIONRESULT

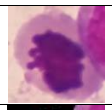
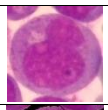
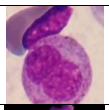
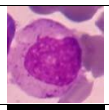
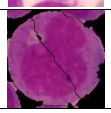
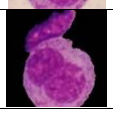
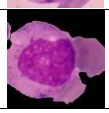
Image Type	Reason			
	Non-WBC Cell	Cell body is cut off	Cell area failed to separate	Cell area exceeds the Body
Original Image				
Segmentation Result				

TABLE 5.4. INCORRECT CELL TERMS AND REASONS

C. FEATURE EXTRACTION RESULTS

All the features of each cell data are stored and act as input data at the classification stage. The average values of cells have been calculated and shown in Table4.5.

Feature	T1	T2	T3	T4	T5	T6	T7	T8
Area	9097.808	15167.818	10404.577	11941.691	12997.013	16690.750	2713.324	2895.563
Nucleus Ratio	0.731	0.562	0.521	0.676	0.685	0.533	0.760	0.768
Perimeter	347.466	465.752	403.381	406.249	423.651	600.205	202.372	192.585
Circularity	0.936	0.876	0.787	0.899	0.897	0.589	0.840	0.920
Mean	126.048	137.439	133.256	120.318	117.407	144.822	103.126	101.208
St. Deviation	16.294	19.780	24.854	17.847	17.384	22.765	21.247	21.698

TABLE V. AVERAGE FEATURE EXTRACTION RESULT OF EACH CELL TYPE

The columns T1 to T8 in Table V stand for the eight successive cell types, they are myeloblast, promyelocyte, granulocyte, monoblast, promonocyte, monocyte, megakaryoblast and support cell. From the average data value of features can be seen characteristics of cell types. For example, monocyte cells have the largest area, whereas myeloblasts have the highest circularity value compared to other cell types.

D. CLASSIFICATION RESULTS

The classification stage begins with the training of data with a model as the output and be used in the testing phase. Based on the test results of 20% of data from each cell type, experiments were conducted on linear and non-linear SVM kernels. Several experiments also performed by combination of features obtained from Minimal Overlap Probability calculations. The parameters used in the training and data testing phase are the best C (cost) and y (gamma) values. From the experimental results on feature combinations, RBF kernel is getting better results than the linear kernel, but in the final results with the highest accuracy average values obtained by the linear kernel that includes all of the features. Details of each accuracy on each feature combination with the linear kernel and RBF is shown in Fig. 5.2.

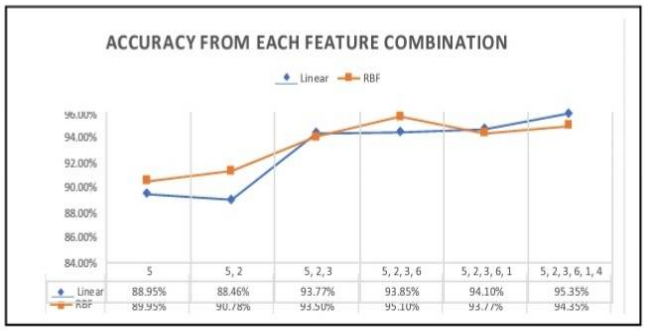


Figure 5.2 accuracy results for each feature combination

Furthermore, the accuracy of each type of cell is selected based on kernel type with the highest accuracy. The kernel that obtains the highest accuracy is linear kernel that includes all combinations of the six extracted features. The detailed of the highest accuracy of each cell obtained by the linear kernel is shown in Fig 5.3.

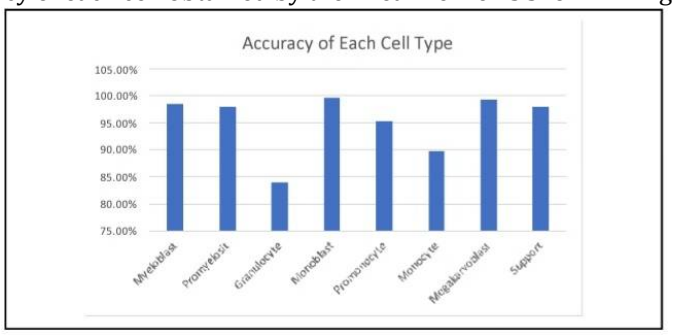


Figure 5.3 results of classification accuracy of each cell type

VI. CONCLUSIONS

Segmentation technique using K-Means with WDT and morphological reconstruction operation can be well applied to segment the bone marrow blood cell object. It proved by the result of an average segmentation of 87.72% from 1710 total real cells. The presentation of correct segmented object data from each AML M4, M5 and M7 preparations were 87.98%, 76.97% and 95.55%, respectively. The supervised classification with multiclass support vector machine with one-vs-rest model has also been applied with good results. The best kernel that can separate the eight data types based on the experiment is linear kernel. The best result of it includes all feature data with the accuracy value of each cell types of 98.67%, 98.01%, 84.05%, 99.67%,95.35%, 89.70%, 99.34% and 98.01% respectively.

ACKNOWLEDGMENT

We are thankful to the management of Sengunthar College of Engineering, Tiruchengode, Tamil Nadu for providing all the amenities to complete this work successfully.

REFERENCES

1. Leukimia and Lymphoma Society, "The AML Guide Information for Patients and Caregivers", 1311 Mamaroneck Avenue, Suite 310, Whute Plains, New York, 2012.
2. American Cancer Society, "Colorectal Cancer Facts & Figures 2014- 2016". Atlanta, Ga: American Cancer Society,2014.
3. Dacie and Lewis, "Practical Haematology, Eleventh Edition", Elsevier Churcill Livingstone, London, ISBN-13: 9780702034084,2011.
4. E. Suryani, Wiharto, and N. Polvonov, "Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia", *International Journal of Computer Science & Network Solutions*, nol. 2, no. 6, pp.35–49,2014.

5. M. Madhukar, S. Agaian, and A.T. Chronopoulos, "Deterministic Model for Acute Myelogenous Leukemia Classification", *International Conference on Systems,* pp.433–438,2012.
6. L. Putzu and C. 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.
7. L. H. Nee, M. Y. Mashor, and R. Hassan, "White Blood Cell Segmentation for Acute Leukemia Bone Marrow Images," *Journal of Medical Imaging and Health Informatics*, vol. 2, no. 3, pp. 278–284, 2012.
8. A. A. Nasir, M. Y. Mashor, and R. Hassan, "Classification of acute leukaemia cells using multilayer perceptron and simplified fuzzy ARTMAP neural networks," *International Arab Journal of Information Technology,* vol. 10, no. 4, pp. 356–364,2013.
9. M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K Means Clustering and Support Vector Machine Classifier," *Journal of Medical Signal and Sensors* , vol. 5, no. 1, pp. 49–58,2015.
10. D. Goutam and S. Sailaja, "Classification of Acute Myelogenous Leukemia in Blood Microscopic Images using Supervised Classifier," *International Conference on Engineering and Technology (ICETECH),* pp. 3–7,2015.
11. K. Laosai, Jakkrich and Chamnongthai, "Acute Leukemia Classification by Using SVM and K-Means Clustering," *Electr. Eng. Congr. (iEECON), 2014 Int.*, pp. 1–4,2014.
12. K. Anwar, A. Harjoko, and S. Suharto, "Feature Selection Based on Minimum Overlap Probability (MOP) in Identifying Beef and Pork," *International Electrical Engineering Congress (iEECON) 2014,* vol. 7, no. 3, pp. 316–322,2016.
13. N. Pradana T.P, E. Suryani, Wiharto, "Pemanfaatan Seed Region Growing Segmentation dan Momentum Backpropagation Neural Network UntukKlasifikasiJenisSelDarahPutih", *Seminar Nasional Teknoin 2013 FTI UII*, Yogyakarta,2013
14. **Dr.R.Satish Kumar** and Dr.K.Umadevi " Novel Technique for Measurements of Dielectric Properties and Microwave Heating of In-Shell Eggs without Explosions in Microwave Oven for Pasteurization", International journal of Innovative research in Advanced Engineering, Vol. 2, , 2015, pp. 69-77 (Impact factor 1.311).

