International Journal of Pure and Applied Mathematics
Volume 118 No. 20 2018, 3731-3737
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

INVESTIGATION OF HEMATOLOGICAL DISORDERS FROM BLOOD CELLS USING SOFT COMPUTING TECHNIQUES: A REVIEW

Aldrin karunharan. K
Research scholar: Department of Electronics and communication
Shri Jagadhishprasad Jhabarmal Tibrewala University
aldrin@imco.edu.om

Om Prakash
Professor: Department of Electronics and communication
Shri Jagadhishprasad Jhabarmal Tibrewala University
om4096@gmail.com

Abstract—Leukocytes are the cells of immune system derived in the bone marrow as hematopoietic stem cell. Thus they act as the battling cells. The presence of immature cells changes the granularity and geometry of leukocytes. Feature extraction is extraction of a vector in multidimensional space from an image, where each dimension represents the attribute of the image that is believed to carry information that is useful in classification of the image. Features such as nucleus and cytoplasm area, average color co-ordinates and number of pixels in the nuclear perimeter are used. Accurate classification of human blood cells plays a decisive role in the diagnosis and treatment of diseases. Hematological disorders refer to the diseases caused with the changes in blood cells or blood system such as Leukemia, Anemia, Malaria and Azotemia. This review paper explores the techniques used in the automatic detection of such hematological diseases which can e recognized using various algorithms.

Keywords—segmentation, lymphoblast, feature extraction & clustering

1.INTRODUCTION

White blood cell composition of the blood gives valuable information and plays an important role in the diagnosis of different diseases like Leukemia. Leukemia is a rapid proliferation of abnormal white blood cells (leukocytes) Anemia can be detected the reduction in the morphology of red blood cells (erythrocytes). Thalassemia, a type of malaria is due to the abnormality in red blood cells [6]

Azotemia is another type of hematological disorder caused by excessive nitrogen compounds which can be detected by the change in micro morphological features from blood cell. Proposed research work begins with the segmentation of white blood cells from blood plasma using new advanced computing methods and new segmentation algorithms [1]. New and already done (by the scholar) and some other existing segmentation algorithms will be compared for their accuracies [10], to determine the optimal algorithm finally to be adopted for the next stage of work.

After the segmentation of leukocytes soft computing classifiers are used to classify using the extracted morphological indexes as Neutrophils, Basophils, Eosinophils, Lymphocytes and Monocytes. Fuzzy C means clustering (FCM) will be used for the classification of pixels which is an unsupervised fuzzy classification algorithm. Separation of the classes can be done using three most relevant features such as cell and nucleus area and the gray intensity of the cytoplasm.

The malarial parasitic detector can use a Bayesian pixel classifier to mark the stained and unstained classes of pixels using conditional probability density functions. Early identification of leukemia, malarial parasites and other hematological disorders can greatly increase the probability of recovery [13]. Blast cells, for instance are characteristics of a certain type of leukemia and would indicate further tests if found in blood. Features are to be extracted according to an efficiency criterion on the basis of classification or recognition tasks.

Being able to automatically able to classify these and flag samples accordingly could be a great boon to hematologists. This would of course require leukemic blood with unusual evidence to be available and manual classification by hematologists for the training data set. A set of special domain features that are extracted will be used to classify whether the tumor is acute myeloid d leukemia or acute lymphoblastic leukemia.

Features such as area, perimeter, convex area, solidity, major axis length, orientation filled area, ratio between cell and nucleus area, mean gray level, rectangularity and circularity can be used for the classification of leukocytes. Fighting with the infections become very difficult when the nature of white blood cells is abnormal. Frequent infections, easy bruising, bleeding and fatigue are the symptoms of leukemia.

We have to find the origin of leukemia whether it is from myeloid cell or lymphoid cell. Leukemia is characterized by the predominance of highly immature blast cells and increased

number of more mature cells. The table 1 given below shows the source of blood diseases and their origin.

Cells of origin	Disease	Pathology
Erythrocyte	Increased RBC	Polycythemia
	Decreased RBC	Anemia
Leukocyte	Increased WBC	Eosinophilia
	Decreased WBC	Sepsis
Hemostatic	Platelet disorder	Thrombocythemeia
	Coagulation disorder	Hemophilia
	Vascular disorder	Purpura

Table 1 Hematological disorders

The classifier must achieve lowest miss-classification error with lowest standard deviation. Being able to automatically able to classify these and flag samples accordingly could be a great boon to hematologists. This would of course require leukemic blood with unusual evidence to be available and manual classification by hematologists for the training data set.

A set of special domain features that are extracted will be used to classify whether the tumor is acute myeloid d leukemia or acute lymphoblastic leukemia. Features such as area, perimeter, convex area, solidity, major axis length, orientation filled area, ratio between cell and nucleus area, mean gray level, rectangularity and circularity can be used for the classification of leukocytes. The classifier must achieve lowest miss-classification error with lowest standard deviation.

Furthermore, the corresponding processing time must not be very high as compared with other tested classifiers [4]. In this research, a methodology will be proposed to detect acute lymphoblastic leukemia by selecting the lymphocyte cells from blood image that is interested by the acute leukemia [9] using morphological indexes from those cells thereby classifying the presence of leukemia [2].

The extraction of this index can be done by iterative erosion filtering of the binary nucleus. The correct number of lobes can be easily extracted from a number of connected objects with areas larger than the fixed threshold. For examples, Neutrophils have 5 lobes, Eosinophils have 2 lobes, Basophils have 3 lobes and Lymphocytes will have always 1 lobe. The goal is to establish programming tools for the identification of different white blood cell categories of a given blood sample using the spatial domain features extracted from the blood cell image and provide information to physicians or hematologists in the form of diagnostic reference to the specific state of acute lymphoblastic leukemia [3] or acute myeloid leukemia (AML) [9], thereby to improve the classification accuracy with the use of soft computing techniques such as artificial neural networks(ANN), Swarm Intelligence and other Evolutionary Computing techniques and then evaluating their performances.

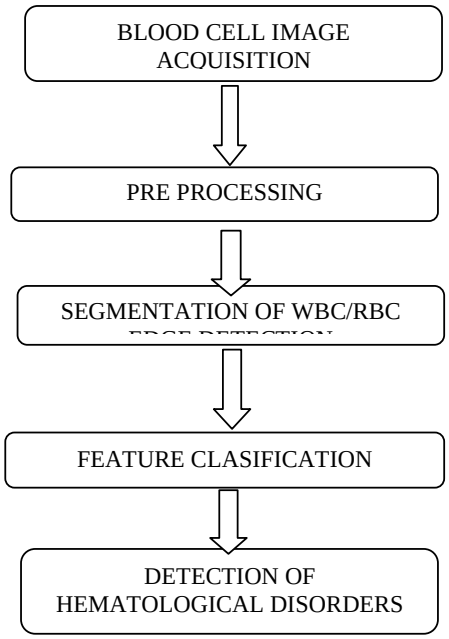

Figure 1: Flow diagram of the detection process

The overall aim of the research project is to investigate and recognize hematological disorders such as acute lymphoblastic leukemia [5], acute myeloid leukemia, anemia and malarial Thalassemia [7] by applying soft computing methods such as artificial neural networks (ANN), Swarm Intelligence and Evolutionary Computing and then evaluating their performances based on their classification accuracy.

2. DETECTION OF MALARIAL PARASITES

Malaria, a deadliest disease can be manually diagnosed, by the screening of blood samples of patients using light microscopes. There is probability of error during manual diagnosis and moreover it is time consuming process. Alternate technique to the manual process is the automated diagnosis of Malaria using computer vision that can be accurate and faster. The first step is to separate the image into regions of interest called segmentation Regions of interest in this case are single red blood cells. For example, while detecting leukemia the region of interest the white blood cells.

Conventional light microscope is used for examining the Gisma stained peripheral blood cells in the microscopic diagnosis of malaria. Methanol is used to preserve the cells where one drop of blood is spread over a single layer called stained structure. A robust method should be used for preprocessing of the images and the calibration of data acquisition set up. Field selection is another challenge because a single blood slide could have thousands of separate field at the required magnification.

We need to automate the selection of fields that are to be analyzed since blood is not uniformly distributed among the fields. Based on the criteria such as thickness and number of cells an algorithm needs to be developed which is fully controllable to decide the fitness of the fields.

The algorithms that developed in the paper can process the images regardless of the density of the field. Blood cells in thin film can be partitioned and labeled for further process. But thick films will not have resolvable RBC. Degree of infection is proportional to the ratio of number of infected red blood cells to the number of healthy cells. Prior to the examination blood sample is always stained to aid the observation and detection of parasites. Staining process highlights the parasites, WBC's and platelets. Definition of the parasites is required prior to classification. Automatic detection algorithms must have the following considerations

i. Unsupervised classification algorithm can be applied to the whole of microscopic slide image for the detection of RBC no matter whether the cells are infected or healthy.
ii. Within the RBC, the detected pixels may be grouped and the location of cells may be determined
iii. The number of RBC in the image can be determined/counted by the grouped objects.
iv. A second unsupervised algorithm may be applied to detect whether these RBCs are infected by malarial parasites or healthy.

Initial attempts of detection of parasites was done using geometrical shapes which resulted in two disadvantages.

a. Overlapping of neighboring RBC which may be seen under segmented
b. Time consuming computation while detecting large no RBC's

I. TEMPLATE FITTING TECHNIQUE

a. Perceptual constraints are used to choose hypothesis for the under to segmented regions
b. For grouping the hypothesis fitting energies E(i) need to be compared
c. To remerge the regions indicated by global consistency, recursive splitting is used
d. Detection rates and mean of template fitting costs E(i) may be evaluated

II. INDEX TREE TECHNIQUE

a. Two index trees are eventually used. One is used for image segmentation and the other is used for retrieving by reorganizing the segmentation tree and both need to be computed offline.
b. Understate changes to the weightage used in the fitting and constancy energies

$$E(i) = \alpha E_{color} + (1 - \alpha)(1 - \beta) E_{Area} + \beta_{Deformed}$$

c. Segmentation index tree is constructed in the boundary deformation parameter space using the cluster based normalized central moment shape features
d. To preserve accuracy, a mapping from feature space of central moments to E(i) the template fitting cost is used
e. A query system is developed to retrieve the shapes of particular interest using the histograms of shape deforming parameters
f. RBC's are clustered and classified according to their shape and by the precision recall curves. They are compared with the previous methods of retrieval based on color histograms and the integration of color and information

III. ATTRIBUTE OPENING TECHNIQUE

Attribute opening technique [AOT] is an advanced morphological operation technique based on binary gray scale images, computes the opening of a gray-scale image for which the gray-scale difference is the highest amongst all possible openings. It is very important to decide how this morphological light microscopy images are converted into gray scale and color space which is the average of RGB channels.

Mathematical morphology is the study of shapes or connected components on binary or gray-scale images. Techniques from mathematical morphology have been used in medical image analysis and especially in automated malaria diagnosis. With mathematical morphology it is possible to _alter elements from an image that fulfill certain attributes. Morphological Openings and Closings are effective tools for isolating arbitrary, parameterized shapes. Basic morphological operations, erosion, dilation, opening and closing, use structuring elements that alter the structure of an image during altering. More advanced morphological _techniques can avoid this problem, and therefore are much more suitable for segmentation purposes.

Researches who worked on segmentation of RBC has proposed to use morphological filters to analyze the size

distribution of the elements on a gray-scale image of red blood cells which actually compute the granulometry, or pattern spectrum using connected set openings for a circular structuring element with increasing radius. This estimated average size of the granulometry peaks of the red blood cells can be used to improve the segmentation, as the size of the regions we are looking for is known. Area granulometry has gained a lot of recognition in research on automated Malaria diagnosis. Researchers also used area granulometry to estimate the average size of red blood cells on an image. They also used area granulometry as a size-estimator in their study. Moreover, they proposed different techniques for actual segmentation of images.

Parameter	I [Intensity]	R[Red]	B[Blue]	G[Green]
Area under ROC	0.9875	0.9789	0.9678	0.9769
True positive rate[TPR]	0.9675	0.9567	0.9768	0.9683
False positive rate [FPR]	0.0345	0.365	0.2967	0.456
Error rate [e]	0.0678	0.0654	0.0567	0.0765

Table1: Performance of the classification shown for intensity red blue and green respectively for various threshold

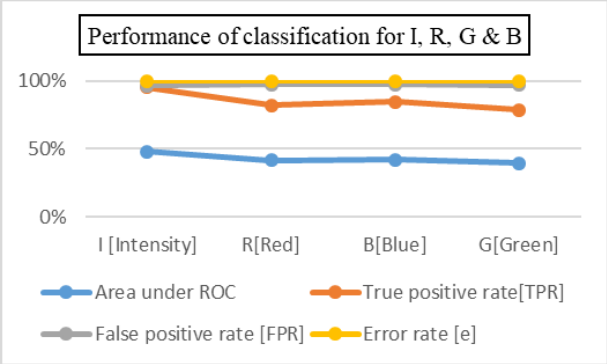

Fig 2: Performance of classification for I, R, G & B

The supervised approaches use fine-tuned parameters such as thresholds and back tracking operations which correct errors which inevitably occurred. Moreover, unsupervised approach needs less number of image processing steps and does not require tuning of parameters & extensive training sets. It is important to determine whether the pixels are from plasma (WBC and body of RBC) or from malarial parasites. This is called discrimination which is unsupervised and the pixel labels are assigned according to certain criteria. Machine learning of such parameters and thresholds are not required.
The Otsu algorithm is based on Gaussian mixture models where color discrimination is carried out using the clustered pixels in color space. The segmentation is done by the set of parameter estimates and pixel labelling or non-parametric statistical criterion which reflects the overall properties of pixel data. Using discrimination approach one searching classification of pixels which maximizes the class variance. Pixels are classified which belongs to RBC or plasma respectively. Pixels are classified as belonging fore ground or background such as to RBC and plasma respectively based on minimized and maximized threshold. The threshold p(T) will vanish in the segmentation process at the gap of the histogram. The starting value of the threshold is chosen equal to the mean such that it converges within few iterations. This algorithm is data driven unsupervised procedure where threshold T is varied in the region of interest from low to high value and the classifier is least mean square type.
If the spatial structure is very variable then the detection of infection level malarial infections and its stage can be easily done by the sequential, hierarchical application of multidimensional extension of the Otsu algorithm. The algorithm is sensitive to the color contrast and RBC density. Normally pattern recognition techniques depend son underlying models and unsupervised approaches do not require training.

3. DETECTION OF LEUKEMIA

Conventional microscopic analysis is time consuming and tedious. We have developed computer aided techniques and algorithms for automatic recognition of acute lymphoblastic leukemia by the use of machine learning where quantitative method for blood sample analysis is very much required. For the detection of leukemia segmentation is necessary to separate mature lymphocyte and malignant lymphoblast cell into constituent morphological regions. We have used both supervised and unsupervised frameworks to address the problems in segmentation which includes feature space clustering, rand1omfield modeling and neural networks. Formulations such as pixel classifications, pixel clustering and pixel labelling are used for our segmentation purpose. Achieving highest classification accuracy is our main goal. Features such as morphological, color and texture are extracted from the nucleus and cytoplasm of lymphocyte images.

The steps in the methodology includes segmentation of lymphoblastic image, nucleus and cytoplasm extraction and classification. Cancer is a deadly disease that can develop in any organ, tissue blood, nerve, lymph node, bone, breast and skin and it is identified by the invasive and uncontrolled growth. In that round 20% of cancer is of hematologic origin. These malignancies can develop from myeloid and lymphoid cell which named as acute lymphoblastic leukemia [ALL] and acute myeloid leukemia respectively. The diseases from

myeloid line are myelodysplastic syndromes and myelogenous leukemia and lymphomas, lymphocytic leukemia, and myeloma are from lymphoid origin.

Because the symptoms of leukemia are similar to other diseases, it is very difficult to detect leukemia in early stages leading to high mortality rate. Blood cancer is otherwise called liquid cancer that develops from blood cells, none marrow and lymphatic system. It is unique from other cancers as it does not produce any mass or tumors. The abnormal white blood cells flood the bone marrow encroaching the space for red blood cells and platelets. This condition of over white blood cells leads to the inability of fighting against infections and the decrease in red blood cells can cause anemia. More over the clotting ability is reduced due to the decrease in platelets.

Fighting with the infections become very difficult when the nature of white blood cells is abnormal. Frequent infections, easy bruising, bleeding and fatigue are the symptoms of leukemia. We have to find the origin of leukemia whether it is from myeloid cell or lymphoid cell. Leukemia is characterized by the predominance of highly immature blast cells and increased number of more mature cells. Prior to cell detection a suitable segmentation scheme to separate white blood cells has to be used. Henceforth image segmentation is the first key step for the detection of leukemia. It is difficult to select optimum threshold in histogram based methods because deep valleys of histogram cannot be located properly.

The list of segmentation algorithms is given below.

i. Watershed method
ii. Deformable models
iii. Clustering/Classification
iv. Histogram–based thresholding

The contours are evolved by internal and external forces to find the region of interest in the robust segmentation approaches. In the deformable models nucleus contour must be initially identified before the segmentation. It is clear that clustering oriented techniques are suitable f1or segmentation than histogram based thresholding methods. In intensity based segmentation, the accuracy of segmentation is low and pixel contextual information is not utilized for the calculation of final pixel class label. Each pixel has three color attributes and is represented by a feature vector. The existing methods are capable of detecting only blast cells even though various researchers has proposed different methods for AML and ALL classification. It can be also observed from the literature that most the schemes rely on extraction of nucleus and very few schemes are able to extract and segment successfully with robustness and high degree of accuracy.

Because of the direct use of gray level intensity or RGB the segmentation of cytoplasm resulted in poor accuracy. More over these feature are linearly inseparable in the image plane. Many schemes have failed to classify boundary pixels such as nucleuses to cytoplasm and cytoplasm – background due to overlapping of RGB sub regions. Due to these factors, it can be said the differentiating the blast between ALL and AML cells are difficult. In a nutshell we can say that there is lot of scope to improve the schemes suggest by the pre researchers who worked in this domain. The objective is to improvise segmentation schemes by utilizing morphological, color and texture features. Final goal is to classify the lymphocytes and lymphoblasts based on WHO criteria.

There by we tend to create an automatic learning system for the classification of leukemic blasts in peripheral blood smear from lymphoid and myeloid origin. Detection of the immature lymphoblasts begins with the initial screening of peripheral blood smear samples. Image processing techniques and machine learning methods are used to differentiate lymphocytes from lymphoblasts. All can be diagnosed by blast counting which is based on lymphocyte image analysis. To facilitate that analysis, the blood cell image has to be segmented to individual morphological regions such as nucleus, cytoplasm and background. The set of pixels in each region which is segmented by must possess identical set of attributes and properties. This results in the homogeneous regions with unique label. The efficacy of all the schemes used has to be validated by the hematologists.

The desired region of interest [ROI] should at least contain one single lymphocyte to aid the detection of ALL. In order to differentiate lymphoblast from mature lymphocyte it is necessary to evaluate the smear image. K means clustering technique is helpful to obtain nucleus image as cluster output using RBG features. We got different cluster outputs for very different run of the k clustering algorithm because of the random initialization of the center. Therefor e each cluster that belongs to nucleus image is represented by average intensity value.

The average of the coordinates of each pixel is used to find the centroid which is the binary representation of the nucleus. The rectangular sub image cropped after the coordinates of centroid is found and each sub image will contain at least one single lymphocyte. The color space normally consists of two layers namely the luminosity layer and chromisity layer denoted as *a and b* and hence the three color RGB is transformed into two that is *a* and *b*. however the gray scale version of lymphocyte image can also be used instead color for the initial input. Illumination variance is corrected using preprocessing technics such as wiener filtering.

Set *a*	Set *b*	Label	Description
0.2475	0.4573	R1	Cytoplasm
0.3456	0.4845	R1	Cytoplasm
0.9863	0.2863	R2	Nucleus
0.9678	0.2876	R2	Nucleus
0.1987	0.9644	R3	Background
0.1897	0.9456	R3	Background

Table 2: Training pattern for sample *a* and *b*

Hence the color space is separated into three regions cytoplasm back ground and nucleus. The classifier is a single structure. According the hematologist each stained lymphocyte image is considered to consist of three regions and hence the Leishman stained images will also be three. Feature extraction and supervised region labeling are used.

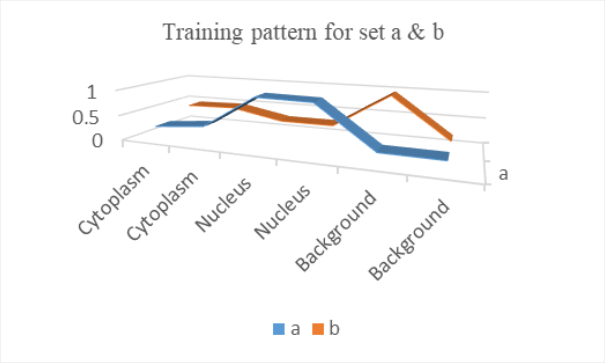

Fig 3 Training pattern for sample set a & b

Pixel selection is done by human expert and graphical tools. Accordingly input- output patterns are generated for training the FLANN with around 300 sub images by sing just 20 sample images. This includes images from benign [lymphocyte] and malignant [lymphoblast]. The color features for each band are easily accessible and are useful in class labelling.

4. CONCLUSION

In this review paper, the blood samples that are collected from oncologists and hematologists were formatted and preprocessed using various denoising and edge detection algorithms. We found that Gabor filter and sober edge operators gave better results. We have analyzed the key features that has to extracted for the investigation of hematological disorders. Moreover, various segmentation algorithms that are applied to segment erythrocytes for detecting malarial parasites and leukocytes for detecting leukemia are compared by data extracted from the research articles. It was observing that clustering method is optimal compared to other methods. We shall be proceeding to find the best techniques for the detection of both malarial parasites and leukemia.

REFERENCES

1. Wang Shitong and Wang Ming , "A New Detection Algorithm based on Fuzzy Cellular Neural Networks for white blood cell detection" *IEEE transactions on Information Technology in Biomedicine* ,VOL 10 , January 2006.
2. Farag.A,"Computer based Acute leukemia classification " , Proceedings of the *46th IEEE International symposiums on Circuits and Systems*, December 2003.
3. Fabio Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope",Proceedings of CIMSA 2005 – *IEEE International Conference of Computational Intelligence for Measurement Systems and Applications* Giardini Naxos, Italy, 20-22 July 2005.
4. Vincenzo Piuri and Fabio Scotti, "Morphological classification of blood leukocytes using Artificial Neural Networks " Proceedings of the *IEEE International conference on Computational Intelligence* , July 2004.
5. Nuannuan Zong, Melvin Ayala and Malek Adjouadi, "Artificial neural networks approaches for multidimensional classification of acute lymphoblastic leukemia gene expression samples" Proceedings of the *9th WSEAS International Conference on Computing* , 2005
6. Toha S.F. Ngah, "Computer Aided Medical Diagnosis for the Identification of Malaria Parasites" Proceedings of International conference on *Signals, Networking & Communications*, Feb-2007.
7. F.Boray Tek1, Andrew G. Dempster and Izzet Kale1,"Malaria Parasite Detection in Peripheral Blood Images ", IEEE *Transactions on Pattern Analysis & Machine Intelligence*, 2005.
8. Sheikh and Michelle,"Blood cell Identification using Neural Networks ", Proceedings of the *IEEE International conference on Bioengineering*, March 1996.