GLRLM based Feature Extraction for Acute Lymphoblastic Leukemia(ALL) Detection

Sonali Mishra, Banshidhar Majhi, Pankaj Kumar Sa, and Faiza Siddiqui

Pattern Recognition Research Lab
Dept. of Computer Science and Engineering
National Institute of Technology, Rourkela, India, 769008
smishra.nitrkl@gmail.com,{bmajhi,pankajksa}@nitrkl.ac.in,faiza.sid16@gmail.com

Abstract. This paper proposes a gray level run length matrix (GLRLM) based feature extraction technique for the detection of Acute Lymphoblastic Leukemia (ALL). ALL could be a fatal hematopoietic ailment which might cause death if it's not treated at the early stage. The GLRL matrix is a method for extraction of statistical textural features from the nucleus of the lymphocyte image. The extracted features are then supplied to the Support Vector Machine (SVM) for classification. The experiments are performed on an publicly available dataset ALL-IDB1. The accuracy of the proposed scheme is found to be 96.97% for SVM classifier.
Keywords. Acute Lymphoblastic Leukemia, Grey level run length, marker-based watershed segmentation, CAD system

1 Introduction

The process of counting and grouping of blood cells from peripheral blood smear allow evaluation and diagnosis of a huge number of diseases. Illness related to hematopoietic cells influence the blood and bone marrow and also are major concerns for death [1]. By analyzing white blood cells (WBCs) or leukocytes, leukemia is usually detected. Based on the rate of progression of the disease, leukemia is categorized into two types i.e. acute and chronic. Acute lymphoblastic leukemia (ALL) is a subtype of acute leukemia which primarily affects the lymphocyte (a type of WBC). One of the demonstrative strategies incorporates the microscopic examination of the white blood cells with abnormalities. From decades, this operation is performed by the experts and skilled operators that suffers from several disadvantages like slowness and non-standard accuracy. Image processing techniques can be a way which provides information on the morphology of the cells. The primary objective of the research is to contribute a fully automated way to support medical activity by analyzing microscopic images.

The remainder of the paper is arranged as follows. Section 2 shows some of the valuable works for the discovery of the disease followed by the proposed system model in Section 3. Section 4 shows the experiments performed along with the comparison made with some standard classifiers. Finally, Section 5 presents the conclusion .

2 Related work

According to the survey, some of the existing systems can analyze and classify the leukocytes form the peripheral blood smear. However, these systems are

partially automatic. Especially, the work has been performed so as to count the number of WBCs through segmentation. Madhloom *et al.* [2] have proposed an automatic system based on arithmetic and threshold operations for segmenting the lymphoblast cells. In [3], the authors have extracted a single leukocyte by applying bounding box around the nucleus. The authors in [4] have used a low-pass filter and a thresholding technique to segment the white blood cell by removing the background. Halim *et al.* [5] have suggested a method for detection of acute leukemia in blood that detects leukocytes by examining the S component of the HSV color space. We can conclude from literature survey that automation of the process depends entirely on the right segmentation and extraction of specific features. In this paper, a segmentation method proposed by [7] along with a new scheme is used for computerized investigation of the blood sample.

3 Proposed Work

The overall block diagram for the detection of ALL is given in Figure 1. The step wise description is given in the subsequent subsections.

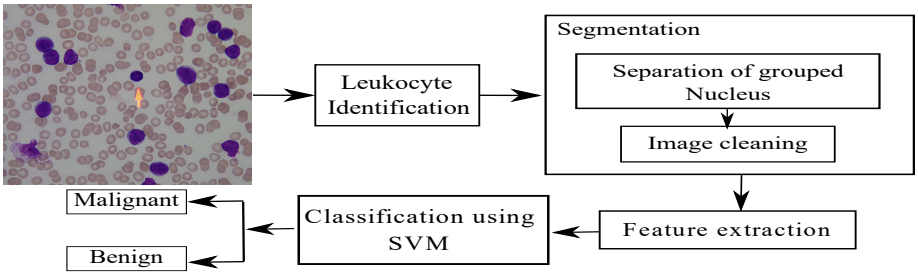

Fig. 1. Proposed block diagram for the detection of ALL

3.1 Nucleus identification

The images made employing a digital microscope is in RGB color space. Color images are very difficult to segment. Therefore, RGB images are converted to CIELAB color space which helps in reducing the color dimension. Here, we have considered Otsu method [8] of thresholding technique for identification of leukocytes from the microscopic image. Figure 2 (a-c) represents the steps for detection of leukocytes.

3.2 Identification and separation of grouped nucleus

The resulting image from the previous step contains only leukocytes. In this step, we have dealt with the separation of grouped leukocytes. This work analyses the presence of grouped leukocytes by taking the roundness value. Roundness value of a cell can be defined by,

$$roundness = \frac{4 \times \pi \times area}{convex\_perimeter^2} \tag{1}$$

In this work, we have considered a roundness value of 0.8 to distinguish between a single leukocyte and a grouped leukocytes. A marker-based watershed [7] is used to refine the line of leukocytes having an irregular shape. The result for the separation of grouped leukocytes is given in Figure 2 (d).

3.3 Image cleaning

The next step after separation of grouped leukocytes is to clean the image. Image cleaning considers the elimination of all the components discovered at the edge of the smear. Cleaning operation can be performed by calculating the solidity value. Solidity of an object is defined as,

$$solidity = \frac{area}{convex\_area} \tag{2}$$

Here, we have taken the solidity value as, 0.85 which is the threshold. So all the objects having less than 0.85 threshold values are excluded from the images. Figure 3 gives the final results after performing the cleaning operation. The individual nucleus can be found out using bounding box technique. Figure 4 shows the details of the sub-imaging process and the corresponding nucleus sub-image.

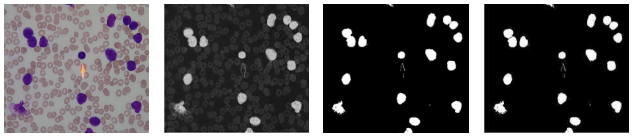

Fig. 2. Nucleus Identification: (a). original image, (b). a^* component of the CIELab color space, (c). Threshold image, (d). Separation of adjacent leukocytes using marker-based watershed segmentation

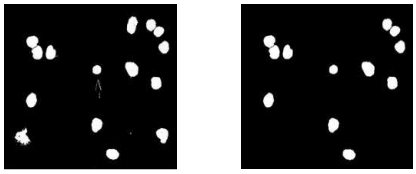

Fig. 3. (a) Image after edge cleaning, (b)Image after removing abnormal component

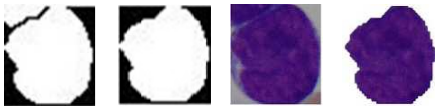

Fig. 4. Sub-imaging process: (a). nucleus sub-image, (b). Image after border cleaning, (c, d). corresponding color image of a and b

3.4 GLRLM based Feature extraction

Many methods of texture analysis have been developed over the past decades [6]. Xiaoou Tang [9] has introduced 11 textural features calculated from the gray level run length matrices which are used to characterize the nucleus of leukocytes. For a sub-image of size $M \times N$, the number of gray levels and the longest run (a string of continuous pixels having the same gray level intensity in a specific linear direction) is represented as g and r respectively. Z is denoted as the total number of run. The GLRLM is a two-dimensional matrix of ($g \times r$) components in which each component $q(m, n)$ gives the times of occurrences

of the run having length n of gray level m in a given direction θ. A run-length matrix $q(m, n)$ is defined as,

$$q_{mn} = |m^n| \tag{3}$$

where, m^n means m exhibits exactly n times, and $1 \leq m \leq g$ and $1 \leq n \leq r$. The feature matrix is calculated using GLRLM and the steps are is described in Algorithm 1.

Algorithm 1 GLRLM based Feature Extraction

Require: Samples of n leukocytes
$GLRLM$: Gray level run-length matrix
θ: direction parameter taken as 0°, 45°, 90°, and 135°
p: number of directions (4)
s and m denotes the feature descriptor and number of features respectively.
Ensure: $X[n : m]$: Feature matrix,
Initialize the value of s
$m \leftarrow p \times s$
for $i \leftarrow 1$ to n **do**
  Calculate the GLRL matrix using $grayrlmatrix()$ for the input image (IP)
  for $q \leftarrow 1$ to p **do**
    $GLRLM_{\theta_q} \leftarrow graycomatrix(IP, \theta_q)$
    for $x \leftarrow 1$ to s **do**
compute the $GLRLM_{\theta_q}$ and append it to X
    end for
  end for
end for

Table 1. Calculated GLRL features

Features	0°		45°		90°		135°	
	healthy	unhealthy	healthy	unhealthy	healthy	unhealthy	healthy	unhealthy
SRE	0.32	0.0186	0.33	0.01	0.32	0.018	0.33	0.018
LRE	29.59	60.50	30.510	60.78	0.32	60.46	30.47	60.78
LGRE	50.82	25.81	47.67	7.71	29.60	26.11	49.91	7.77
HGRE	180.95	34.09	275.003	62.46	49.81	34.29	276.77	62.35
SRLGE	0.564	0.20	0.89	0.36	177.005	0.19	0.894	0.36
SRHGE	0.099	0.04	0.132	0.11	0.53	0.04	0.12	0.11
LRLGE	3107.31	762.38	1127.42	252.64	0.097	743.68	1029.46	253.11
LRHGE	0.0127	0.0012	0.020	0.002	3294.28	0.001	0.018	0.002
GLNU	2439.16	12.87	880.90	4.40	0.011	12.62	782.4	4.40
RLNU	2.36	1.89	3.81	6.42	2613.5	1.86	3.79	6.45
RPC	32350.99	47631.85	10424.5	15644.35	2.44	0.018	10242.68	15663.13

3.5 SVM based classification

SVM is a binary classifier which generates a hyperplane by employing a subset of training vectors which are known as support vectors. This paper uses Support Vector Machine [10] for the classification process. To evaluate the effectiveness of SVM model, the proposed method is being compared to many standard models, namely, k-NN (k-Nearest Neighbour), Naive Bayes, Back Propagation Neural Network (BPNN). Along with this, the proposed method is being tested with the most common kernel used in SVM.

Table 2. Performance evaluation different classifiers over 5-fold

Classifier	True positive rate (TPR)	True negative rate (TNR)	Accuracy(%)
NB	0.99	0.99	96.27
k-NN	0.73	0.85	83.49
BPNN	0.81	0.99	95.36
SVM-L	**0.87**	**1.00**	**96.97**
SVM-Q	0.88	0.99	96.60
SVM-P	0.95	0.65	73.21
SVM-R	0.86	0.96	96.51

Table 3. Comparison of Accuracy (%) with the other existing scheme

Classifier	90 features [12]	Proposed Method (44 features)
Naive Bayes	81.66	96.27
k-NN	83.46	83.27
BPNN	58.7	95.36
SVM-L	**89.76**	**96.97**

4 Experimental evaluation

The experiment is being carried out using MATLAB R2015b on Microsoft Windows 8.1 having a 4 GB RAM as internal memory. The proposed method is trained and tested with the public database ALL-IDB1 [11]. A 5-fold cross validation scheme is used to generalize the performance of the classifier. The lymphoblasts (malignant), and leukocytes(normal) are termed as positive and negative class respectively. This paper takes into consideration of true positive rate (TPR), true negative rate (TNR), and accuracy as performance measures which are defined as follows,

$$TPR = \frac{TP}{TP+FN}, \quad TNR = \frac{TN}{TN+FP}, \quad Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

where, TP= true positive, FP=false positive, TN=true negative, FN= false negative

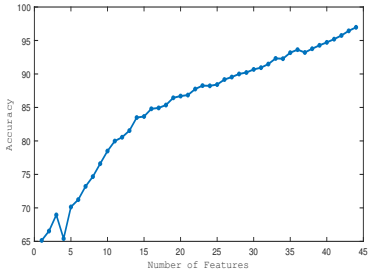

Fig. 5. Plot of accuracy with the increase number of features

4.1 Results and discussion

ALL is diagnosed by the appearance or lack of unhealthy leukocytes samples. Therefore, leukocytes must be characterized as "unhealthy" or "healthy" cells in blood samples for the diagnosis of ALL. So, a total number of 865 individual nuclei were obtained by the sub-imaging process. After finding number of nucleus, the next step is to extract the texture features using GLRL matrix from the sample images. The total number features extracted along all the directions is 44. The texture features extracted from the nucleus regions is tabulated in Table 1. Finally the result of the feature extraction gives us feature matrix of size

865×44. The feature matrix is split into two parts. The first part of size 435×44 is used as training data and the rest of size 430×44 is used for testing purpose. Table 2 presents the performance evaluation of the recommended system. Experiments show that the maximum accuracy of 96.97% has been obtained using SVM classifier along with the linear kernel. Figure 5 shows the performance of the suggested method varying the features number. Finally, Table 3 describes the relative analysis of the suggested scheme with the existing scheme.

5 Conclusion

In this work, we have suggested a method of leukocyte identification by designing an automatic system which is more reliable than the work done by operators manually and is computationally less expensive. The proposed method can efficiently detect the leukocytes present in a blood smear and can classify lymphoblast with great precision, as indicated by our results. The proposed method incorporates the GLRL features to classify the lymphoblast cells and shows an accuracy of 96.97% with SVM classifier.

References

1. Mishra, S., Majhi, B. and Sa, P.K., A survey on automated diagnosis on the detection of Leukemia: A hematological disorder. In 3rd IEEE International Conference on Recent Advances in Information Technology (RAIT), pp. 460-466, 2016.
2. Madhloom, H.T., Kareem, S.A. and Ariffin, H., An image processing application for the localization and segmentation of lymphoblast cell using peripheral blood images. Journal of medical systems, 36(4), pp. 2149-2158, 2012.
3. Mohapatra, S., Patra, D. and Satpathy, S., An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. Neural Computing and Applications, 24(7-8), pp.1887-1904, 2014.
4. Scotti, F., Robust segmentation and measurements techniques of white cells in blood microscope images. In Proceedings of the IEEE Conference on Instrumentation and Measurement Technology (IMTC), pp. 43-48, 2006.
5. Halim, N.H.A., Mashor, M.Y. and Hassan, R., Automatic blasts counting for acute leukemia based on blood samples. International Journal of Research and Reviews in Computer Science, 2(4), 2011.
6. Galloway, M.M., Texture analysis using gray level run lengths. Computer graphics and image processing, 4(2), pp.172-179, 1975.
7. Mishra, S., Majhi, B., Sa, P.K. and Sharma, L., Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection. Biomedical Signal Processing and Control, 33, pp.272-280, 2017.
8. Otsu, N., A threshold selection method from gray-level histograms. Automatica, 11(285-296), pp.23-27, 1975.
9. Tang, X., Texture information in run-length matrices. IEEE transactions on image processing, 7(11), pp.1602-1609, 1998.
10. Bishop, C.M., Pattern recognition. Machine Learning, 128, pp.1-58, 2006.
11. ALL-IDB Dataset for ALL Classification. `http://crema.di.unimi.it/~fscotti/all/`
12. Mishra, S., Sharma, L., Majhi, B. and Sa, P.K., Microscopic Image Classification Using DCT for the Detection of Acute Lymphoblastic Leukemia (ALL). In Proceedings of International Conference on Computer Vision and Image Processing (CVIP), pp. 171-180, 2017.