Biomedical Signal Processing and Control 47 (2019) 303–311

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection

Sonali Mishra*, Banshidhar Majhi, Pankaj Kumar Sa

Pattern Recognition Laboratory, Department of Computer Science and Engineering, National Institute of Technology, Rourkela 769008, India

ARTICLE INFO

Article history:
Received 8 June 2017
Received in revised form 21 March 2018
Accepted 8 August 2018

Keywords:
Acute lymphoblastic leukemia (ALL)
Discrete orthonormal S-transform (DOST)
AdaBoost with RF (ADBRF)
Computer-aided diagnosis (CAD)

ABSTRACT

This paper presents an effective scheme for classification of the normal white blood cells from the affected cells in a microscopic image. The proposed method initially pre-processes the input images using Y component of the CMYK image and a triangle method of thresholding. Subsequently, it utilizes discrete orthonormal *S*-transform (DOST) to extract the texture features, and its dimensionality is reduced using linear discriminant analysis. The reduced features are then supplied to the proposed Adaboost algorithm with RF (ADBRF) classifier where the random forest is used as the base classifier. A publicly available dataset, ALL-IDB1 is used to validate the proposed scheme. The simulation results based on the five runs of *k*-fold stratified cross-validation indicate that the proposed method yields superior accuracy (99.66%) as compared to existing schemes.

1. Introduction

The visual examination of blood samples is a major criterion for the analysis of leukemia [1,2]. Acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) are the two different types of Leukemia which can lead to death if not treated at the right time. AML affects the myeloid organs, whereas ALL is seen in the bone marrow. ALL is a significant hematopoietic disease which is generated by the abnormal collection of white blood cells (WBCs). With the increase in the number of malignant WBCs, the fighting capability of the body with the foreign material gets diminished. The early detection of ALL can considerably improve the probability of recovery, especially in the case of children [3,4]. The recognition of blast(unhealthy WBCs) cell in the bone marrow is also an important step for the detection of ALL. The percentage of blasts is a major concern for detecting the proper stage of the ALL and is also helpful in the proper treatment of the patients. According to French-American-British standard (FAB) [5], three distinct types of ALL are characterized based on the morphological differences among the lymphoblast [6].

So far, the detection of the disease highly depends upon the perfection of the hematologists and pathologists. To support the hematologists, a computer-aided diagnosis (CAD) is a basic necessity for accurate classification and early detection of ALL. The main step in a CAD system is to generate features of WBCs which will classify the cells as healthy or affected. The most identifiable feature of a normal blood cell can be categorized as morphological, statistical, and textural features. Further, it is also necessary to classify different blast types. A typical blood smear having a lymphocyte (healthy) and a lymphoblast (affected) are given in Figure 1. In this paper, we have proposed a texture based feature in microscopic images using Discrete Orthonormal *S*-Transform (DOST). The suggested feature defines the characteristic of an image as rough, smooth, silky, or bumpy as a function of the spatial variation in pixel intensities.

Rest of the article is organized as follows. Section 2 describes the related work. Section 3 outlines the proposed method for the detection of ALL. Section 4 presents the details about the data source and gives a relative comparison of the proposed methodology. Finally, the conclusion is given in Section 5.

2. Related work

In recent years, many researchers have been working on the development of CAD systems. Investigations have been made for the detection of lymphoblasts cells in microscopic images. They have taken into consideration for various morphological, textural, and color features for the detection of the disease. Those features are then classified using different classifiers. The computer-assisted discovery and analysis techniques can be broadly divided into two categories. The first category applies the genetic information, while

* Corresponding author.
E-mail address: sonaliabc.mishra@gmail.com (S. Mishra).

https://doi.org/10.1016/j.bspc.2018.08.012
1746-8094/

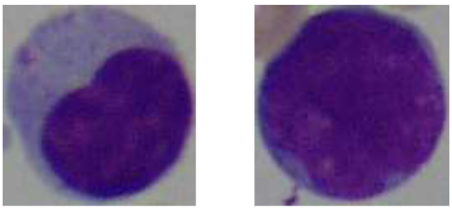

Fig. 1. Representation of a peripheral blood smear having a lymphocyte and a lymphoblast.

the second one uses the information present in the image modeled by different machine learning techniques.

Lin et al. [7] have suggested an approach for prediction of tumor in the microscopic blood images. They have used genetic algorithms for feature selection and silhouette statistics to differentiate between six subtypes of ALL. They have shown a better accuracy rate by using microarray data and gene expression. The testing accuracy using 23 genes are found to be 100%, and CFS/SVM only performed an accuracy of 96% with the help of more than 20 predictive genes. Zong et al. [8] have proposed a technique for the detection of lymphoblasts based on flow cytometer data and reported a classification accuracy of 96.67%. But it is very difficult to extract the gene expression from the bone marrow samples, and it requires very sophisticated equipment. Ross et al. [9] have suggested an scheme to differentiate different types of pediatric ALL. They have analyzed blasts from 132 samples using higher density nucleotide arrays. A set of newly selected genes is incorporated into class predicting algorithms to get an overall accuracy of 97%.

Escalante et al. [10] have suggested an application to the problem of acute leukemia detection. They have classified the subtypes of acute leukemia using an ensemble particle swarm optimization selection technique (EPSMS). The classification of acute leukemia based on morphological features can be done by building ensembles. For a 2 class classification, the authors have got an accuracy of 97.68% whereas the accuracy got decreased to 94.21% for multiclass classification. Foran et al. [11] have proposed a framework for differentiating lymphoma and leukemia. They have found an accuracy of 83% on 19 different cases of leukemia and lymphoma. Though they have differentiated leukemia and lymphoma, the suggested method has not been validated in ALL and on a larger dataset. Scotti et al. [12] have suggested distinguishing different WBCs or leukocyte by examining the morphological properties of a color image. The proposed system first distinguishes between the WBC and other components of blood cells. This procedure provides satisfying results in finding different components which show a way for identification of tumor deformation in the cell morphology. The authors have taken a dataset of 134 images containing leukocytes. For differentiating different types of WBCs, they have used parallel FF-NN to get an accuracy of 92% with a feature size of 23. In another work, Scotti [13] suggested a scheme for detection of ALL from the microscopic images. The experiments are being conducted on 150 images and shown that morphological features are more feasible for lymphoblast recognition for the detection of ALL with a classification error 0f 0.0133 using feed-forward neural network.

Halim et al. [14] have presented an approach to count the number of blasts in the case of ALL. They have taken histogram based thresholding technique succeeded by S-component on the HSV space. Subsequently, morphological erosion is performed for counting the blast cells. The overall accuracy of the proposed system is found to be 97.8% with a very small dataset consisting of 50 images. Also, the authors have not specified the threshold value used for separating nucleus and cytoplasm. Mohapatra et al. [15] have recommended an ensemble of classifier system in which accuracy has been improved by analyzing morphological and textural features from the peripheral blood smear having an accuracy of 99% with ALL-IDB1 dataset. Putzu et al. [16] have proposed an approach which isolates the whole leukocyte from a microscopic image and subsequently separates the nucleus and cytoplasm. For every cell, distinct features like shape, color, and texture are extracted and are used to train different classification models to determine the best one for leukemia classification. The authors have found the accuracy of 93.2% with the help of 131 features. Angulo et al. [17] utilized watershed segmentation for lymphocyte identification, where morphological properties are extracted for characterizing lymphocytes in light of cell typology. Though this technique shows exact results for segmentation, it has not been used for classification.

It has been observed that morphological, textural, and color-based features are predominant while classifying lymphoblasts. Among classifiers, ANN and SVM have been widely used. In this paper, we have proposed a lymphoblast classification scheme using features extracted from discrete orthonormal S-transform (DOST) followed by feature reduction using a hybrid approach which includes PCA+LDA. Finally, the discriminant features so obtained are passed to Adaboost Random Forest classifier. The proposed model uses 2D discrete orthonormal S-transform (2D-DOST) as the feature extractor. The 2D-DOST is based on a set of orthonormal basis function that preserves both time-frequency and phase information of a signal. Thus, DOST gives features with zero information redundancy.

3. Proposed work

The proposed methodology has different phases like any other classification scheme which include the preprocessing, sub imaging, feature extraction, feature reduction, and classification. We have utilized the standard preprocessing techniques like noise reduction and smoothing of background. The main contribution lies on DOST based feature extraction and PCA+LDA based feature reduction. The relevant features are subsequently used for classification on an AdaBoost based random forest (ADBRF) classifier. The block diagram of the proposed scheme is depicted in Figure 2. The phases are discussed below in sequel.

3.1. Pre-processing, sub-imaging and PBS segmentation

The images from the ALL-IDB1 dataset have been collected under different magnification and as a result, comprised of noise and background effects. The RBCs and platelets present in the smear are unwanted for the detection of disease. To extract the WBCs from the blood smear, background subtraction is performed. Since the *Y* component of the image contains maximum information regarding the WBC, the original RGB image is regenerated to CMYK color space. Triangle method [18] for thresholding has been used for extracting the WBCs from the background. The overall steps followed in pre-processing are described in Figure 3.

Microscopic blood images are relatively larger in size and consists of more than one WBCs per image. However, the region of interest must contain only one WBC for the detection of ALL. Marker-based watershed segmentation [19] is used to separate the grouped cells. The use of marker-based watershed segmentation results in separating the grouped leukocytes by imposing the markers on the blood smear. The result for marker-based watershed segmentation is depicted in Figure 4. The images taken for this experiment is a combination of normal and abnormal images present in the ALL-IDB1 dataset. A more detailed explanation of the proposed marker-based watershed segmentation scheme can be found in our previous work [19].Though all the WBCs require