IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

NCAIT 2017

JSS Academy of Technical Education

Vol. 4, Special Issue 8, May 2017

Effective Method for Leukocytes Classification and Segmentation in Blood Smear Images

Mahitha G[1], Nikshepa T[2], Spoorthy K P[3]

Assistant Professor, Dept. of ISE, DSATM, Bengaluru, India[1,2]

Student, Dept. of ISE, DSATM, Bengaluru, India[3]

Abstract: Blood is one of the important component of the body. It consists of RBC, WBC and Platelets. Detection and counting of white blood cells (WBC) in blood samples provides valuable information to hematologists, to identify various types of hematic pathologies such as AIDS and blood cancer (Leukemia). But performing this task manually prone to error and time consuming. An automatic detection and classification of WBC images can enhance the accuracy and speed up the detection of WBCs. In this paper, we propose an efficient framework for localization of WBCs within microscopic blood smear images using a multi-class ensemble classification mechanism. In the proposed framework, the nuclei are first segmented, followed by extraction of features such as texture, statistical, and wavelet features. Finally, the detected WBCs are classified into five classes including basophil, eosinophil, neutrophil, lymphocyte, and monocyte. The proposed method improves the segmentation performance when compared to other state-of-the-art segmentation methods.

Keywords: Hematology; Image Segmentation; Image Classification;multi-class ensemble.

I. INTRODUCTION

In the last two decades, a number of automatic and semiautomatic methods are proposed for the segmentation and classification of medical diagnostic modalities. It has been observed in medical field that majority of the diseases in the body can be identified by analyzing blood samples [1]. There are numerous categories of blood cells such as red blood cells (erythrocytes), white blood cells (leukocytes), and platelets (thrombocytes). Furthermore, each of these blood categories are further divided into various classes. The total count of blood cells and fractional count of WBCs provide important information to doctors for identification and diagnosing of different kinds of diseases. To accomplish this task, there are two possible methods: Manual segmentation of nucleus from WBCs and their classification based on a set of parameters, which is inherently difficult,prone to errors, and time consuming due to involvement of humans. Furthermore, the instruments used by experts for manual segmentation and classification of WBCs are not affordable by all hospital and clinics. To avoid these problems, automatic techniques can be used for detection and classification of WBCs based on different machine learning algorithms. This is evident from different medical imaging softwares, which automatically diagnose various types of diseases using WBCs [2].

WBCs consist of five sub-categories known as monocyte, lymphocytes, basophile, eosinophil, and neutrophil. In order to diagnose, correctly detect WBC, and its underlying sub-class, a multiclass classification is considered as the best option, which can be used to efficiently classify each category. Image classification is based on different image features like HOG, edges, geometric, texture, and statistical features based on which different images are compared and classified [3]. The first step in image classification problem is preprocessing that includes image sharpening, contrast adjustment, and noise removal. Different techniques are used for the enhancement of microscopic images. The enhanced image is further processed for segmentation of WBCs using different segmentation techniques such as manual thresholding [4], OTSU binarization, fuzzy C-mean [5], and active contours [6]. Active contours are well known and widely used in various applications for medical image analysis.Fuzzy C-means, (FCM) algorithm is another unsupervised clustering technique used in image segmentation, allowing a piece of data to belong to two or more clusters [7].

Our proposed method uses a popular unsupervised machine learning technique known as color k-means clustering algorithm. K-means is an automatic segmentation algorithm whose speed depends on the number of clusters K. According to this approach, similar intensities are clustered in the same cluster while different intensities are clustered to other clusters based on the value of K, which is selected manually. The proposed framework segments the WBCs into four clusters. Firstly, the enhanced RGB image is converted to HSI color, model. Next, color k-mean is used to segment the WBC from the image. Next, features from the segmented WBC are classified using support vector machine (SVM). The extracted features include: 1) statistical features such as mean, variance, standard deviation, root mean square (RMS), regression, skewness, and kurtosis, 2) texture features such as correlation, gray scale co-occurrence

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

NCAIT 2017

JSS Academy of Technical Education

Vol. 4, Special Issue 8, May 2017

matrix (GLCM), entropy, energy, and inverse difference moment, which are extracted from 300 different types of WBCs. Our contributions are summarized as follows:

1. A novel computer aided system is proposed for leukocytes classification and segmentation in blood smear images, helping hematic pathologists in diagnosing various diseases more efficiently with better accuracy.
2. For efficient and effective segmentation, color k-means clustering algorithm is incorporated in the proposed framework, providing better segmentation results compared to state-of-the-art schemes. Furthermore, the proposed segmentation algorithm is computationally inexpensive, making it more suitable for segmentation.
3. The proposed system considers both statistical and texture features of blood smear images extracted by using transform domain for classification of leukocytes. Furthermore, an ensemble multi-class classification mechanism is devised to classify leukocytes into five different classes.

The rest of the paper is structured as follows: the proposed work is explained in Section 2. The experimental results are given in Section 3, followed by conclusion in Section 4.

II. PROPOSED FRAMEWORK

The framework is proposed for WBCs segmentation and classification with a three step process including 1) WBC'snuclei segmentation from microscopic blood smear images, 2) features extraction from the segmented nuclei, and 3)classification of leukocytes into their respective five categories using a multi-class ensemble SVM [8]. In this work,segmentation is performed using colour k-mean clustering algorithm. After segmentation, the segmented region is transformed to frequency domain and a set of statistical and textural features are extracted. For robust classification onvarious blood smear images under different lightening and noisy conditions, a multi-class ensemble classification schemeis used to classify the leukocytes into five different classes. The proposed framework has the competency to easily segment and classify WBCs into their corresponding five classes. The schematic representation of the proposed framework is shown in Fig. 1.

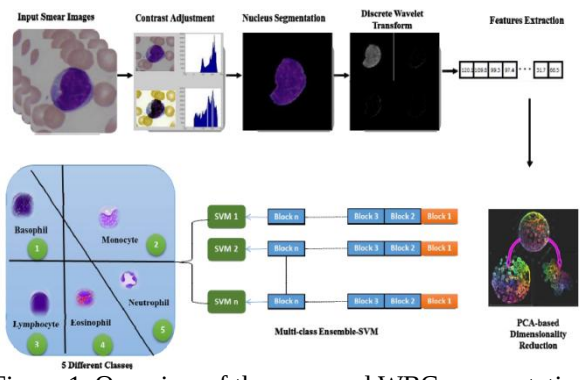

Figure 1. Overview of the proposed WBC segmentation and classification framework

Table I. Image dataset obtained from HMC hospital

Type	Basophil	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Total
No of Images	60	90	350	380	150	1030

A. Data Acquisition and Pre-Processing

The study consists of 1030 blood smear WBC samples which were collected from Hayatabad Medical Complex (HMC) Peshawar, Pakistan. These blood smears were captured with high-definition color camera head Nikon DSFi2. The digital images were taken with approximately 100 magnifications. The details of dataset is given in Table I.The well-known pre-processing includes imageenhancement such as noise removal, contrast adjustment, andimage sharpening. In case of our proposed work, the obtaineddataset does not require much pre-processing as the smearimages in the dataset were collected very carefully. For ourframework, the images are sharped using Gaussian un-sharpmask as sharp images can be easily segmented**.** In theproposed method, the sharper images are converted from RGBto HSI color space for applying colour k-means clusteringalgorithm [9]. The main purpose of converting RGB colourspace to HSI is to minimize the number of colours, helping ineasy segmentation of WBCs using k-mean clustering algorithm [10].

B. Blood Cells Nuclei Segmentation

In computer vision, segmentation refers to division of areas having similar properties. The goal of segmentation is simplifyand change the representation of an image into something more meaningful and easier to analyze [7]. Image segmentation has numerous practical application in different fields especially inmedical imagining such as studying the anatomical structures,diagnosis, treatment planning, localizing tumours, counting leukocytes, classifying WBCs, and other pathologies. Image segmentation partitions an image into a set of disjoint andhomogeneous regions, which are meaningful to a certainapplication [13]. Thus, the segmentation process is based onglobal thresholding, mathematical morphology, fuzzy c-mean clustering, watershed, Otsu binarization, and color contrast. Global thresholding is a good segmentation method for microscopic blood smear images as the cytoplasm, nucleus, and background have their own unique grey levels. This method can perform worst when the lighting level varies from one image to another image. Colour based segmentation of WBCs includes five different techniques to segment them fromother cells of the image. The user can select one of these methods and check the precision and accuracy of different techniques to decide about the correct algorithm for a particular application and disease. In order to select the suitable segmentation method for the proposed framework, we considered K-mean cluster based segmentation, fuzzy c-mean, active contours, Watershed method, OTSO and simple thresholding method. Through experiments, we found color kmeans the best candidate for our proposed framework as it issimple and comparatively more accurate for colourbasedsegmentation.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
NCAIT 2017
JSS Academy of Technical Education
Vol. 4, Special Issue 8, May 2017

C. Features Extraction and Feature Reduction using PCA

After segmentation process, the next step is featureextraction which is one of the critical step towardsclassification accuracy. Features are properties of images,representing their natural similarities. These features alongwith their labels are then used by classifier for matching different images and classifying them into certain classes. In the proposed work, we have extracted three different set offeatures including:

i. Geometric features such as area, perimeter, and centroid.
ii. Statistical features such as arithmetic mean, variance, standard deviation, regression, correlation, skewness,kurtosis, root mean square, and histograms.
iii. Textural features such as energy, entropy, maximum probability, contrast, and inverse difference movement.

These features are extracted using the transform domain (DWT), which is comparatively more suitable to extract strongfeatures for leukocytes classification. To this end, DWT isapplied on each dimension of the 2D blood images, producingfour sub-bands LL, LH, HH, and HL. The process is repeatedtwo more times for LL band. Our proposed method uses level-3 decomposition for feature extraction due to strong features ofLL band at level-3 of DWT. After features extraction, it isnecessary to reduce its size to minimize the computation time and storage requirement. To achieve this goal, a method knownas principle component analysis (PCA) is used. PCA is simpleand more suitable for our framework compared to other dimensionality reduction methods such as auto encoder.

D. Multi-Class Ensemble-SVM for Classification of Segmented Blood Cells

SVM isa well-known supervised learning technique in machine learning, which is based on statistical learning theory. This technique isrobust and accurate even if we have small amount of training101data. In the proposed work, we have used an ensemble multiclassSVM (EMC-SVM) for classification of leukocytes intofive classes. This is due to the diversity of blood smear imagesfor which training a single classifier is impractical because oflimited performance [8],[11]. It has been experimentally provedthat ensemble SVM performs well compared to traditionalSVM [12]. Therefore, the proposed EMC-SVM was devised toclassify the WBCs into five different classes. For trainingpurpose, 75% of the whole data is utilized. The remaining 25%data is used for testing the accuracy of the proposed classifier.To test a new blood smear image, the same procedure of preprocessing, segmentation, and feature extraction is performed.The extracted feature vector is then passed through EMCSVM,which assigns a class label to the given test imageamong the available five classes.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were carried out on white blood smear images collected from HMC. To collect the ground truth data,the Hematology expert was requested to manually classify theWBCs into their corresponding classes, i.e. Neutrophils,Basophils, Eosinophils, Lymphocytes and Monocytes. Thesemanual results were recorded to build up the database toestimate the results of different classification techniques. Theexperiment was conducted on 1030 blood smear images,containing both RBCs and WBCs. From the given dataset ofimages, only the WBCs were segmented and classified intotheir respective classes. Five sub-classes of WBCs are shown in Figure 2. The results were then compared with the groundtruth to calculate the accuracy of the proposed leukocytes classification method. We applied our segmentation method on each of the 1030images and then compared the results with manual segmentation. The accuracy percentages were calculated accordingly. Our dataset obtained an. The segmentation results for single WBC of the proposed method are in Figure 3. The comparison is based on three metrics including recall, precision and f-measure. Classification of leukocytes cells into five subclassesis manually selected by three human users (medicalspecialist) from blood smear under analysis. This labeled data is considered as ground truth and then compared with theresults produced by our proposed multi-class ensemble classifier. The number of matched and unmatched classification results are then calculated.The comparison between ground truth and classification results generated by a technique is used to define the following terms:

True Positive: A leukocyte cell accurately classified into either five categories (basophil, eosinophil, neutrophil, lymphocyte, and monocyte) by both medical specialists and the technique,

False Positive: A leukocyte cell accurately classified into either five categories (basophil, eosinophil, neutrophil, lymphocyte, and monocyte) by technique but not by medical specialist, and

False Negative: A leukocyte cell accurately classified into either five categories (basophil, eosinophil, neutrophil, lymphocyte, and monocyte) by medical specialist but not by technique.

The number of true positive, false positive and false negative frames is used to reflect the quality of the proposed classification method in terms of standard metrics Recall, Precision and F measures defined as

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \quad (1)$$

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \quad (2)$$

$$F = 2.\frac{\mathrm{Re}\,call.\mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \quad (3)$$

The recall, precision and f-measure of the proposed method for five different types of WBCs of the testing data has been stated in Figure 4

The visual quality of the proposed segmentation technique was evaluated over comparison with two state-of-the-art methods, i.e. adaptive threshold and Active contour techniques [6][14].

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

NCAIT 2017

JSS Academy of Technical Education

Vol. 4, Special Issue 8, May 2017

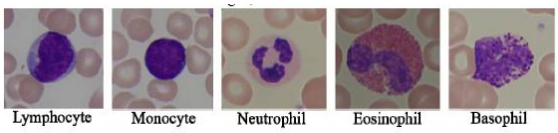

Figure 2. Sample blood smear images, showing five classes of leukocytes cells

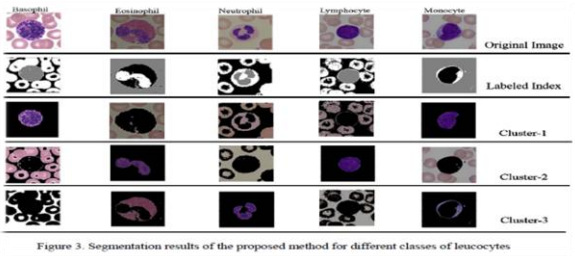

Figure 3. Segmentation results of the proposed method for different classes of leucocytes

For this purpose, the three parameters were considered to check the visual quality of different leucocytes. The parameters are as follows: shape and size. The performance of the proposed method is compared with four state-of-the-art methods as shown in Table II. It can be seen that the proposed method achieves a high overall accuracy for cytoplasm segmentation. Table II shows that the proposed method obtains low false and leak detection ratios and its traced boundaries achieve high accuracy.

IV. CONCLUSION

In this paper, we proposed a framework for classification of leukocytes into their corresponding classes. Firstly, color kmeansalgorithm is used to segment WBCs from blood smear images. Next, morphological operations are performed to the segmented region for removing unwanted components. Then, aset of rich features are extracted from the 3-level DWT decomposition of the segmented region. Based on the extractedfeatures, an ensemble multi-class classification classifier istrained. Through experimental results, we found that the proposed method improves the segmentation performance when compared to other state-of-the art segmentation methods.

Due to the diverse nature of blood smear images, a singleclassifier is almost impractical. Therefore, we proposed anEMC-SVM for classification of leukocytes. Experimentalresults confirmed that the proposed method can successfullysegment WBCs from blood smear images and can classifythem into their respective categories including Neutrophil, Eosinophil, Basophil, Lymphocyte, and monocyte. Theaccuracy of the proposed method was found higher when compared to linear and naïve Bayes classifiers.

Table II. Performance analysis of the proposed method based segmenting each types of WBC in percentage.

Method	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil	Average
Adaptive Threshold-based method [14]	88.7	90.1	92.3	87.2	83.9	92.3
Active-contour-based method [6]	91.9	94.6	94.5	88.3	86.2	89.2
Proposed method	93.6	95.0	98.8	90.4	86.5.7	94.7

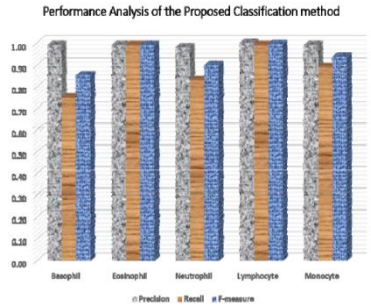

Figure 4. Performance analysis of the proposed ensemble multi-class classification method in terms of precision, recall and f-measure.

ACKNOWLEDGMENT

This research is supported by Sejong University research project “computer aided automated diagnosis system for malaria parasite in thin blood smear images”.

REFERENCES

[1] A. Gautam, and H. Bhadauria, "Classification of white blood cells based on morphological features." Advancesin Computing, Communications and Informatics (ICACCI, 2014 International Conference on. IEEE, 2014.

[2] S. Ravikumar, “Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine,” Artificial cells, nanomedicine, and biotechnology, vol. 44, no. 3, pp. 985-989, 2016.

[3] Z. Liu, J. Liu, X. Xiao et al., “Segmentation of White Blood Cells through Nucleus Mark Watershed Operations and Mean Shift Clustering,” Sensors, vol. 15, no. 9, pp. 22561-22586, 2015.

[4] N. Abbas, D. Mohamad, A. Abdullah et al., “Report: Nuclei segmentation of leukocytes in blood smear digital images,” Pakistan journal of pharmaceutical sciences, vol. 28, no. 5, pp. 1801, 2015.

[5] P. K. Mondal, U. K. Prodhan, M. S. Al Mamunet al., “Segmentation of white blood cells using fuzzy C means segmentation algorithm,” IOSR Jornal of ComputerEngineering, vol. 1, no. 16, pp. 1-5, 2014.

[6] S. H. Rezatofighi, and H. Soltanian-Zadeh, “Automatic recognition of five types of white blood cells in peripheral blood,” Computerized Medical Imaging and Graphics, vol. 35, no. 4, pp. 333-343, 2011.

[7] F. Sadeghian, Z. Seman, A. R. Ramliet al., “A framework for white blood cell segmentation in microscopic blood images using digital image processing,” Biological procedures online, vol. 11, no. 1, pp. 196, 2009.

[8] I. Mehmood, M. Sajjad, and S. W. Baik, “Mobile-cloud assisted video summarization framework for efficient management of remote sensing data generated by wireless capsule sensors,” Sensors, vol. 14, no. 9, pp. 17112-17145, 2014.

[9] J.-f. Li, K.-Q. Wang, and D. Zhang, "A new equation of saturation in RGB-to-HSI conversion for more rapidity of computing." pp. 1493-1497.

[10] J. Duan, and L. Yu, "A WBC segmentation methord based on HSI color space." pp. 629-632.

[11] B. C. Ko, J.-W. Gim, and J.-Y. Nam, “Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake,” Micron, vol. 42, no. 7, pp. 695-705, 2011.

[12] G. Valentini, M. Muselli, and F. Ruffino, “Cancer recognition with bagged ensembles of support vector machines,” Neurocomputing, vol. 56, pp. 461-466, 2004.

[13] J. Prinyakupt, and C. Pluempitiwiriyawej, “Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers,” Biomedical engineering online, vol. 14, no. 1, pp. 1, 2015.

[14] T. N. Pappas, “An adaptive clustering algorithm for image segmentation,” IEEE Transactions on signal processing, vol. 40, no. 4, pp. 901-914, 1992.