Computer Aided System for Leukocytes Classification and Segmentation in Blood Smear Images

Muhammad Sajjad[1], Siraj Khan[1], Muhammad Shoaib[1], Hazrat Ali[2],Zahoor Jan[1]

[1]Digital Image Processing Laboratory, Islamia College Peshawar, Peshawar,
[2]COMSATS Institute of Information Technology, Abbottabad
muhammad.sajjad@icp.edu.pk
hazratali@ciit.net.pk

Khan Muhammad[3], Irfan Mehmood[4,*]

[3]Digital Contents Research Institute
[4]Department of Computer Science and Engineering
Sejong University, Seoul, Republic of Korea
khan.muhammad.icp@gmail.com
irfan@sejong.ac.kr

Abstract—Detection and counting of white blood cells (WBC) in blood samples provides valuable information to medical specialists, helping them to evaluate a wide range of important hematic pathologies such as AIDS and blood cancer (Leukaemia). However, this task is prone to errors 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. Experimental results on a natural (non-synthetic) benchmark database validate the effectiveness and efficiency of the proposed system in contrast to state-of-the-art schemes.

Keywords— Image Segmentation; Image Classification; Hematology; Medical Image Analysis.

I. INTRODUCTION

In the last two decades, a number of automatic and semi-automatic methods are proposed for the segmentation and classification of medical diagnostic modalities. Medical image processing techniques allow us to segment and classify different types of blood cells, soft tissues, and bones from medical images. 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]. The processing of microscopic blood smear images also help us to segment RBCs, WBCs, platelets, count the number of cells, calculate their size, and normal percentage in the human blood. 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 pre-processing 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. The active contour models segment the objects from an image using curve, which starts around the object and moves toward its inner normal. When it reaches the boundary of the segmented object then it stops moving. From the stopping point, the shape of the object is detected [3]. 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

---

* Corresponding Author

IEEE computer society

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 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 in-expensive, 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's nuclei 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 on various blood smear images under different lightening and noisy conditions, a multi-class ensemble classification scheme is 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.
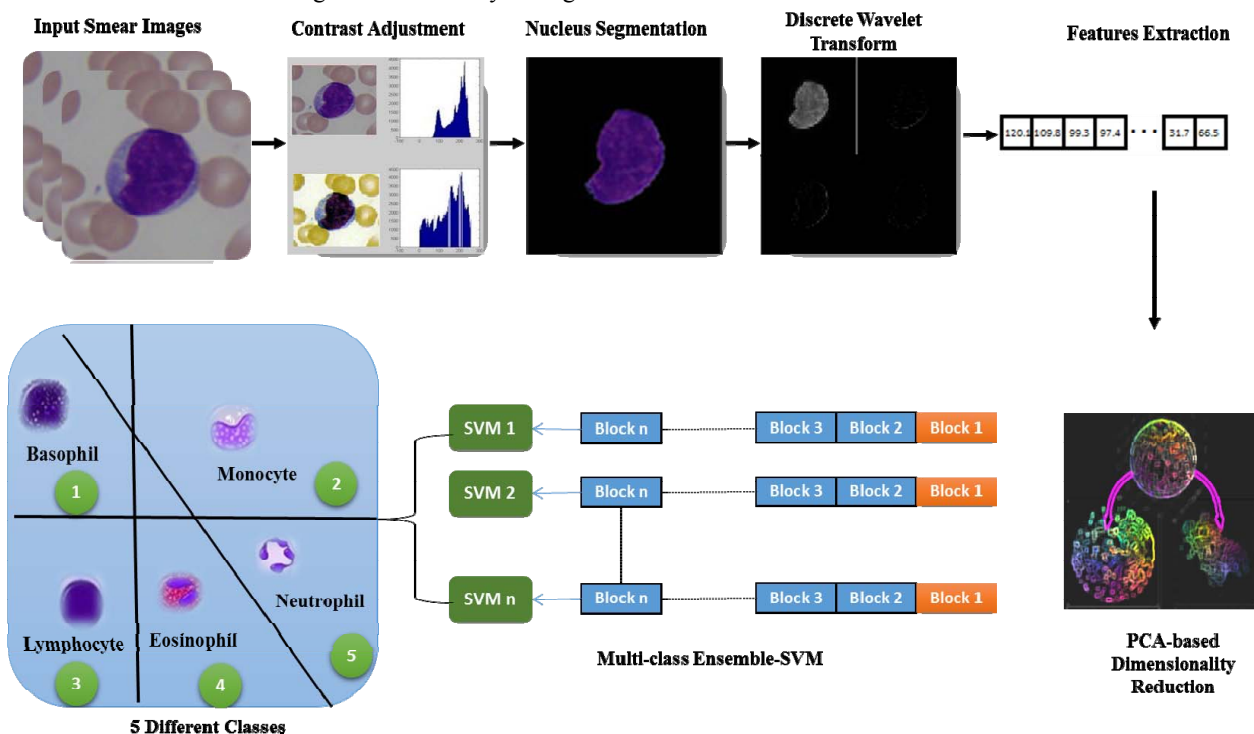

Figure 1. Overview of the proposed WBC segmentation and classification framework

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 [1]) Peshawar, Pakistan. These blood smears were captured with high-definition color camera head Nikon DS-Fi2[2]. The digital images were taken with approximately 100 magnifications. All the images were saved in JPG format of dimension 960×1080 pixels. The sizes of the images were adjusted to 256×256 pixels during processing in MATLB as per the requirement of simulation. The details of dataset is given in Table I.

The well-known pre-processing includes image enhancement such as noise removal, contrast adjustment, and image sharpening. In case of our proposed work, the obtained dataset does not require much pre-processing as the smear images in the dataset were collected very carefully. For our framework, the images are sharped using Gaussian un-sharp mask as sharp images can be easily segmented [17]. In the proposed method, the sharper images are converted from RGB to HSI color space for applying colour k-means clustering algorithm [9]. The main purpose of converting RGB colour space to HSI is to minimize the number of colours, helping in easy 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 simplify and 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 in medical 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 and homogeneous regions, which are meaningful to a certain application [13]. Thus, the segmentation process is based on global 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 from other 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 k-means the best candidate for our proposed framework as it is simple and comparatively more accurate for colour based segmentation.

C. Features Extraction and Feature Reduction using PCA

After segmentation process, the next step is feature extraction which is one of the critical step towards classification accuracy. Features are properties of images, representing their natural similarities. These features along with 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 of features 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 strong features for leukocytes classification. To this end, DWT is applied on each dimension of the 2D blood images, producing four sub-bands LL, LH, HH, and HL. The process is repeated two more times for LL band. Our proposed method uses level-3 decomposition for feature extraction due to strong features of LL band at level-3 of DWT. After features extraction, it is necessary to reduce its size to minimize the computation time and storage requirement. To achieve this goal, a method known as principle component analysis (PCA) is used. PCA is simple and 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

The next step after feature extraction is to select the best classifier, considering the input and its expected result. SVM is a well-known supervised learning technique in machine learning, which is based on statistical learning theory. However, keeping in view the nature of complexity and diversity of the blood smear images, multi-class ensemble SVM is the best feasible solution which can be trained separately for each sub-category of WBCs. This technique is robust and accurate even if we have small amount of training

---

[1] http://www.hmcpeshawar.com.pk/
[2] https://www.nikoninstruments.com/Products/Cameras/Camera-Heads/DS-Fi2/Specifications

data. In the proposed work, we have used an ensemble multi-class SVM (EMC-SVM) for classification of leukocytes into five classes. This is due to the diversity of blood smear images for which training a single classifier is impractical because of limited performance [8, 11]. It has been experimentally proved that ensemble SVM performs well compared to traditional SVM [12]. Therefore, the proposed EMC-SVM was devised to classify the WBCs into five different classes. For training purpose, 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 pre-processing, segmentation, and feature extraction is performed. The extracted feature vector is then passed through EMC-SVM, which assigns a class label to the given test image among 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 the WBCs into their corresponding classes, i.e. Neutrophils, Basophils, Eosinophils, Lymphocytes and Monocytes. These manual results were recorded to build up the database to estimate the results of different classification techniques. The experiment was conducted on 1030 blood smear images,

containing both RBCs and WBCs. From the given dataset of images, only the WBCs were segmented and classified into their respective classes. Five sub-classes of WBCs are shown in Figure 2. The results were then compared with the ground truth to calculate the accuracy of the proposed leukocytes classification method.

We applied our segmentation method on each of the 1030 images and then compared the results with manual segmentation. The accuracy percentages were calculated accordingly. Our dataset obtained an average overall accuracy of 95.7% for nucleus segmentation and 91.3% for cytoplasm segmentation. The segmentation results for single WBC of the proposed method are in Figure 3.

Furthermore, several experiments are conducted to evaluate the performance of the proposed classification scheme. The comparison is based on three metrics including recall, precision and f-measure. Classification of leukocytes cells into five sub-classes is manually selected by three human users (medical specialist) from blood smear under analysis. This labeled data is considered as ground truth and then compared with the results produced by our proposed multi-class ensemble classifier. The number of matched and unmatched classification results are then calculated.
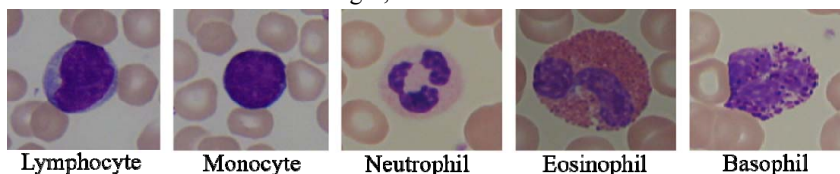

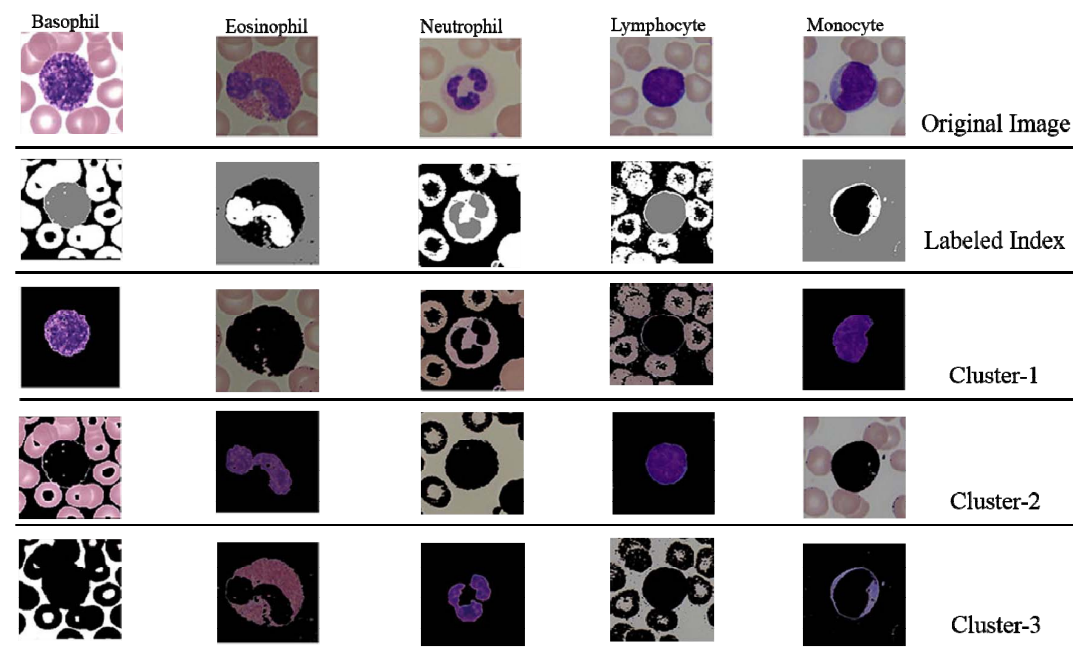Figure 2. Sample blood smear images, showing five classes of leukocytes cells

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

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:

$$\text{Re}\,call = \frac{TP}{TP + FN} \qquad (1)$$

$$\text{Pr}\,ecision = \frac{TP}{TP + FP} \qquad (2)$$

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

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. For testing data, 98.6% precision in average was obtained, which is far better than the averaged accuracy, i.e., 91.8% achieved by [13]. In multi-class problems as in this work, the classification precision, recall and f-measure of neutrophils and lymphocytes are high than other types of WBCs. Hence, the sensitivity (recall) is more appropriate measure for classification of WBC subtypes. The individual sensitivities of each WBC subtype especially minorities like

basophils, eosinophils, and monocytes, were found far better in the proposed framework than other state of the art methods.

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]. 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 k-means algorithm is used to segment WBCs from blood smear images. Next, morphological operations are performed to the segmented region for removing unwanted components. Then, a set of rich features are extracted from the 3-level DWT decomposition of the segmented region. Based on the extracted features, an ensemble multi-class classification classifier is trained. 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 single classifier is almost impractical. Therefore, we proposed an EMC-SVM for classification of leukocytes. Experimental results confirmed that the proposed method can successfully segment WBCs from blood smear images and can classify them into their respective categories including Neutrophil, Eosinophil, Basophil, Lymphocyte, and monocyte. The accuracy of the proposed method was found higher when compared to linear and naïve Bayes classifiers. As a future direction, it would be interesting to explore other late fusion approaches for classification enhancement, such as those discussed in [15], and explore unsupervised representation learning as in [16].

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

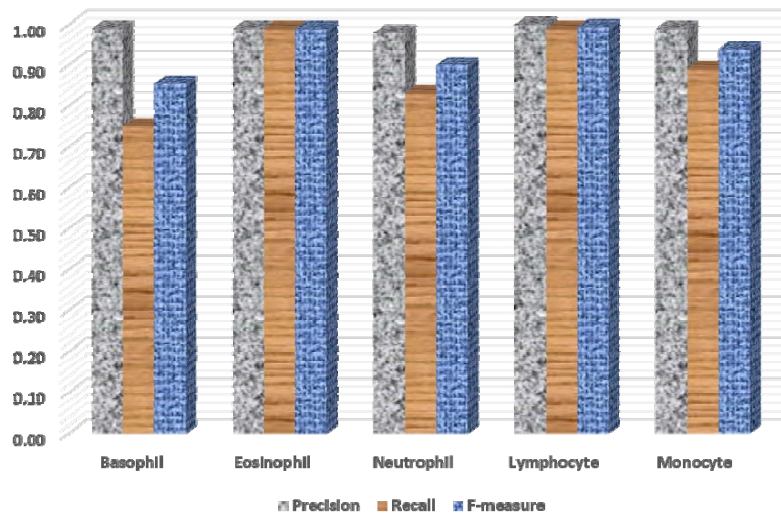| Method	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil	Average
Adaptive Threshold-based method [14]	88.7	90.1	92.3	87.2	83.9	92.3
Active-contour-based method [6]	91.9	94.6	94.5	88.3	86.2	89.2
Proposed method	93.6	95.0	98.8	90.4	86.5.7	94.7

Performance Analysis of the Proposed Classification method

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

REFERENCES

[1] A. Gautam, and H. Bhadauria, "Classification of white blood cells based on morphological features." *Advances in 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 Mamun *et al.*, "Segmentation of white blood cells using fuzzy C means segmentation algorithm," *IOSR Jornal of Computer Engineering,* 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. Ramli *et 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.

[15] H. Ali *et al.,* "Unimodal late fusion for NIST i-vector challenge on speaker detection," *Electron. Lett.*, vol. 50, no. 15, pp. 1098–1100, Jul. 2014

[16] Ali, H. *et al.*, "Speaker recognition with hybrid features from a deep belief network", *Neural Comput & Applic* (2016). doi:10.1007/s00521-016-2501-7