

# Isolation and Classification of Red Blood Cells in Anemic Microscopic Images

Jameela Ali Alkrimi, Loay E. George, Azizah Suliman, Abdul Rahim Ahmad, Karim Al-Jashamy

**Abstract**—Red blood cells (RBCs) are among the most commonly and intensively studied type of blood cells in cell biology. Anemia is a lack of RBCs is characterized by its level compared to the normal hemoglobin level. In this study, a system based image processing methodology was developed to localize and extract RBCs from microscopic images. Also, the machine learning approach is adopted to classify the localized anemic RBCs images. Several textural and geometrical features are calculated for each extracted RBCs. The training set of features was analyzed using principal component analysis (PCA). With the proposed method, RBCs were isolated in 4.3seconds from an image containing 18 to 27 cells. The reasons behind using PCA are its low computation complexity and suitability to find the most discriminating features which can lead to accurate classification decisions. Our classifier algorithm yielded accuracy rates of 100%, 99.99%, and 96.50% for K-nearest neighbor (K-NN) algorithm, support vector machine (SVM), and neural network RBFNN, respectively. Classification was evaluated in highly sensitivity, specificity, and kappa statistical parameters. In conclusion, the classification results were obtained within short time period, and the results became better when PCA was used.

**Keywords**—Red blood cells, pre-processing image algorithms, classification algorithms, principal component analysis PCA, confusion matrix, kappa statistical parameters, ROC.

## I. INTRODUCTION

RED blood cells (RBC) are among the most commonly and intensively studied blood cells in cell biology. The lack of RBCs to the extent "its density is lower than normal hemoglobin level" then this lack case indicates to "anemia" illness.

Anemia is one of the common blood disorders worldwide. It is a public health problem affecting populations in developed and developing countries. Furthermore, the World Health Organization stated that 66% to 80% of the world's population suffers from anemia [1]. Anemia is diagnosed on the basis of the size, color, and shape of RBCs. As such, different types of anemia have been detected [2].

With advancement in technologies, machine learning technique is applied in medical image diagnosis in which computer technology is employed. Image processing is the core of computer vision systems (a part machine learning

applications); it is involved in several stages of the vision system. The typical medical computer vision system consists of four main processing: (i) preprocessing, (ii) localization and segmentation, (iii) feature extraction, (iv) Matching and classification decision.

The main target of pre-processing stage is to manage the visual appearance of microscope images, such that to facilitate the localization and segmentation tasks.

Segmentation is very crucial because the accuracy of the subsequent feature extraction and classification depend on the correct segmentation of the ROI. This step is also among the most difficult parts of automatic cell analysis because considerable uncertainty is observed in microscopic images [3]. In image segmentation, the whole image is divided into unique disjoint regions. The whole process is critical because segmented images should retain, as maximum as possible, useful information and discard irrelevant information [4]. Segmentation is also the most important challenge encountered in previous studies; as such, recent studies have improved this step [5]. Image segmentation techniques can be categorized as supervised or unsupervised. In general, domain knowledge should be combined to solve image segmentation problems effectively [6].

Many image segmentation methods can be found in several literature reviews published in the last three decades. Many general approaches of image segmentation are based on thresholding, cluster analysis, edge detection, region growing, and watershed methods [7]. Various segmentation methods have been presented [8]; however, a gold standard method that can provide good results related to all kinds of microscopic images has yet not been developed [9]. Hence, the objective of this study was to isolate and classify the anemic red blood cells image by three machine learning algorithms.

Features are extracted by transforming image data to quantifiable features, which can be used in the classification stage. The global and local features of images are basic considerations to capture a general visual content from an image. Local features are represented by image spots. Global features can represent the whole image with a single vector [10]. In our study, the selected features are of three main types: color, texture, and geometry.

Principal component analysis (PCA) is a statistical approach used to reduce high-dimensional spectral data and improve classification accuracy of the methods. PCA was first introduced by Pearson in 1901 and independently developed by Hotelling in 1933 [11]. PCA method was transforming the data image into corresponding Eigen value. Three machine learning methods have been used to find the minimum number

J.A. is with the College of Information Technology, University Tenaga National, Malaysia, College of Dentistry, University of Babylon, Iraq (e-mail: jameela\_ali65@yahoo.com).

A.S. and A.A. are with the College of Information Technology, University Tenaga National, Malaysia (e-mail: abdrahim@uniten.edu.my).

L. E. is with the Department of Computer Science, College of Sciences, Baghdad University, Iraq (e-mail: loayedwar57@yahoo.com).

K.A. is with the Faculty of Medicine, SEGI University, Malaysia (e-mail: jashamy@yahoo.com).

of discriminating features required to reach high classification performance.

The objective of this study is to extract the red blood cells areas from the color microscope image, then detect and classify the anemic red blood cells image using three machine learning algorithms.

## II. MATERIALS AND METHODS

In this study 1000 color microscope images have been classified into eight classes; normal and seven types of anemic RBCs. The images are collected from our dataset. The images were collected from 200 hematology slides showing five types of anemic. All of the images medical features have been verified by the staff of the Faculty of Medicine, SEGI University, Malaysia.

As shown in Fig. 1 the established RBC detection and classification system consists of the following stages:

### 1) Loading the Color Image Data

As first stage the data of three color channels (i.e., RGB) are loaded and stored in the computer memory. Among different kinds of images file formats, bitmap file format (BMP) was chosen because of: (i) in contrast to other formats (e.g., TIFF, GIF, and JPG) BMP does not need any encoder, (ii) images data are stored without losing information due to compression (e.g., no distortions) [12].

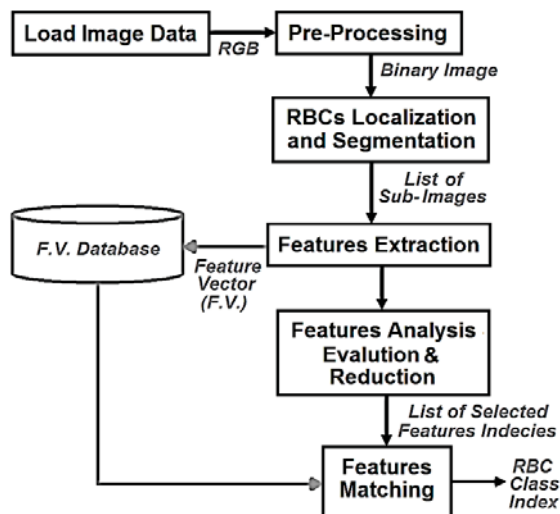


Fig. 1 Isolation and methodology process steps

### 2) Preprocessing

This stage aims to manipulate the color image data to produce a binary map image, such that the white (1) pixels represent ROI points, and black points represent irrelevant (i.e., background) points. As first step the color image data is converted to gray. Then, Gamma mapping mixed with linear stretching is applied, and the output is passed through image binarization based on thresholding. The main reason for applying gamma with linear stretching is to shift the bright area toward white color, and to shift the dark areas toward black. This task will make the process of binarization using

global thresholding is enough and suitable to isolate the RBC from the surrounding areas. Then, the smoothing operation (using mean filter of size 3x3 pixels) is applied twice times. The next step is binary color conversion; it is applied to make the ROI areas as bright areas, and the background as dark area.

To improve the appearance of white ROI (i.e., removing gaps, small holes, and smoothing the ROI ragged boundary) the morphological tiling operation is applied many times. Also, the appeared relatively big black regions (i.e., nucleus area) lay within the ROI white regions are handled using seed filling algorithm to collect them and convert their points color from black to white.

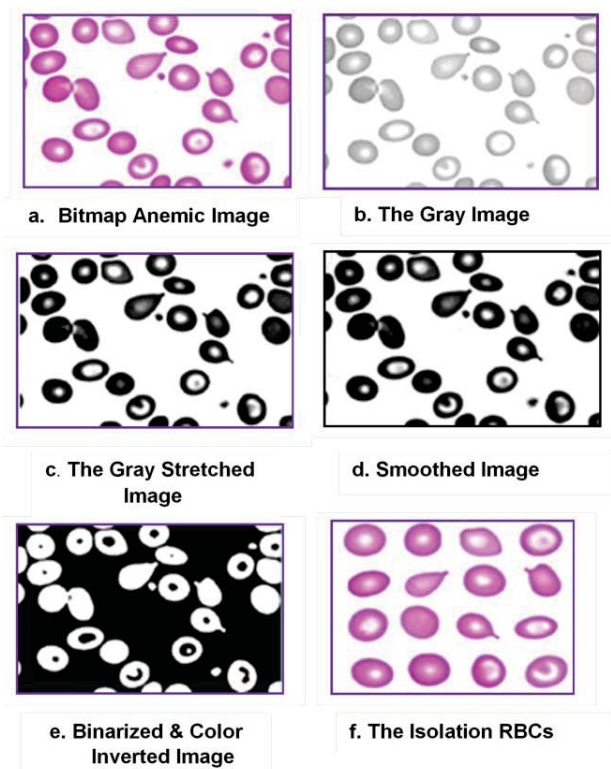


Fig. 2 An illustration of RBCs process

### 3) Localization and Segmentation

These two tasks are accomplished using seed filling algorithm to collect the white pixels of each individual white region. Then, the positions of collected pixels are used to establish color sub-images; each one represents one of the isolated white segments. Each pixel of the sub-image has the same RGB color values of the corresponding pixel in the original image. Fig. 2 present an illustration of RBCs process

### 4) Features Extraction:

In this stage, for each sub image a set of texture and geometrical features are extracted. This set of features holds the following sub-sets of features:

- The histogram moments (up to 5<sup>th</sup> order) for the color components and the corresponding gray image.
- The spatial-moments of the color components and gray values (the taken pivot point is the center of mass).

- c. The contrast histogram moments for each color and gray components.
- d. The gradient histogram moments for each and gray components.
- e. The set of Fourier descriptors for the external boundary.
- f. The aspect ratio of the extracted cell.
- g. The relative size of the nucleus within the Cell.
- h. Some other shading homogeneity measures

The total number of determined features was 271 features.

### 5) Features Analysis and Evaluation

In this stage, the principal component analysis (PCA) algorithm was used to select the features have well is discriminating power. The selected features were then utilized to classify RBCs using one of the three machine learning algorithms. The classification results have been evaluated on the basis of confusion matrix and kappa statistical parameters.

To attain efficient classification of RBCs, we applied inter/intra scattering analysis to filter out the weak features, and keep only the features show low scattering around their centroid values.

### 6) Classification of RBCs

The WEKA platform was used to apply the classification task. WEKA is a free software tool that contains several machine learning algorithms that can be used to solve real-world problems [13]. Three different machine learning algorithms have been applied; namely, radial basis function neural networks (RBFNNs), support vector machine (SVM), and K-nearest neighbor (K-NN) algorithm classifiers. Among the 271 features, a classifier based of seven features was used to classify 1000 different normal and abnormal RBCs taken from 200microscope images. Table I shows the result of each machine learning algorithm.

Three parameters were used to evaluate the classification of RBCs. In particular, confusion matrix evaluated by computing true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results of the tested images [14], kappa statistical parameters, and ROC. These parameters used to verify the suitability of extracted features from RBCs [15]. The metrics were used to estimate the performance of the three classifiers [16] are: (i) Precision (is the percentage of RBC images classified as relevant), (ii) Recall (or sensitivity, is the percentage of relevant RBC images labeled "relevant" by the classifier).

## III. RESULTS AND DISCUSSION

The proposed RBCs isolation method had isolated the individual RBCs within 4.3 seconds (as an average time) from images that showing 18 to 27 cells with high accuracy. Simulations were conducted using the WEKA tool. Three different machine learning algorithms, namely, RBFNN, SVM, and K-NN classifiers, were applied. Fig. 3 shows the attained accuracy results for the three methods. In PCA, the classification time was short and the attained accuracy was high. In RBFNN, the time spent in classification is decreased significantly from 2780 second to 1.47 second when the 7 PC

components are used instead of all feature set. Table IV shows the variation of the classification time between the two cases (all features, only 7 components), for the three algorithms.

Table I shows the evaluation classification result when the 271 features are used. Table II shows the evaluation classification results for PCA when seven features are used. The comparisons between feature and PCA in three methods are presented in Table III. The comparison between the classification accuracy is presented in Fig. 3.

TABLE I  
 CLASSIFICATION RESULTS USING 271 FEATURES

Class. Methods	TP (100%)	FP (100%)	Precision -100%	Recall -100%	Kappa Statistic
SVM	0.97	0.4	0.98	0.97	0.9689
RBFNN	0.965	0.004	0.974	0.965	0.8318
K-NN	0.97	0.3	97	97	0.9689

TABLE II  
 CLASSIFICATION RESULTS USING 7 PCA

Class. Methods	TP (100%)	FP (100%)	Precision -100%	Recall -100%	Kappa Statistic
SVM	0.999	0.009	0.999	0.999	0.9944
RBFNN	0.98	0.3	0.98	0.98	0.9751
K-NN	1	0	1	1	1

TABLE III  
 A COMPARISON BETWEEN FEATURE AND PCA IN THREE METHODS

Class. Methods	271 Feature Class.	7 PCA Class.
SVM	97%	99.90%
RBFNN	96.50%	98%
K-NN	97%	100%

TABLE IV  
 THE CLASSIFICATION TIME OF THE THREE METHODS

Classification	Time (second)	
	Feature	PCA
SVM	7.88	6.89
ANN	2780	1.47
K-NN	2.1	0.01

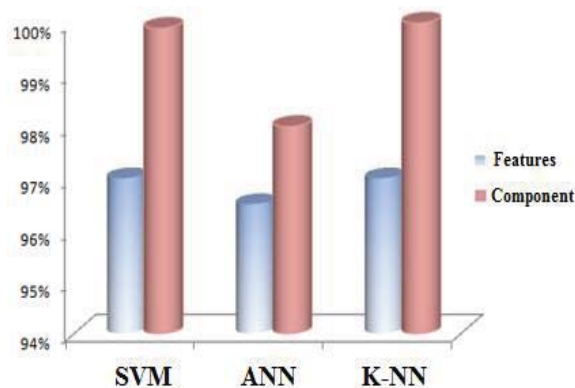


Fig. 3 The classification ratio of features and component

#### IV. CONCLUSION

This study presented two major contributions: a system used to automatically isolate RBCs and a classification method used to distinguish anemic RBCs in microscopic images. The proposed software allowed a highly accurate isolation of individual RBCs in 4.3 seconds for an image containing 18 to 27 cells. From image dataset an efficient features were extracted from the isolated RBCs sub-images. By using machine learning algorithms, we have classified eight types of anemic RBCs. The proposed algorithm involved the following stages: digital image capture; cell isolation; cell segmentation; feature extraction; feature selection by PCA; and classification by WEKA. In K-NN and SVM classifiers, classification results were obtained for a short time period, indicating that classification was improved in terms of accuracy, sensitivity, specificity, and kappa. K-NN yielded 100% of these parameters; SVM resulted in 99.9% accuracy, 99.9% sensitivity, 99.9% specificity, and 96.89% kappa compared with RBFNN. Among the evaluated machine learning algorithms, SVM and KNN classifiers could be used to improve conventional classification methods significantly, particularly those utilized to classify anemic RBCs in microscopic images. On the basis of the viewpoint of an end-user, we found that the results of this study could facilitate laboratory experiments by reducing time and costs.

- [15] Lavesson, Niklas, "Evaluation and Analysis of Supervised Learning Algorithms and Classifiers", Blekinge Institute of Technology, 2006
- [16] Rajendran, P. and M. Madheswaran, "An Improved Brain Image Classification Technique with Mining and Shape Prior Segmentation Procedure", Journal of Medical Systems, Vol. 36, No. 2, Pp. 747-764, (2012).

#### REFERENCES

- [1] Organization (WHO), World Health. "Global Database on Anemia". The database on Anemia includes data by country on prevalence of anemia and mean hemoglobin concentration, (2010).
- [2] Turgeon, Mary Louise, "Clinical Hematology: Theory and Procedures", Lippincott Williams & Wilkins, P. 100, (2012).
- [3] Suzuki, Kenji. "A Review of Computer-Aided Diagnosis in Thoracic and Colonic Imaging". Quantitative imaging in medicine and surgery, Vol. 2, No. 3, P. 163, (2012).
- [4] Adollah, R., Mashor, M.Y., Mohd Nasir, N.F., Rosline, H., Mahsin, H., Adilah, H. "Blood Cell Image Segmentation: A Review", In 4th Kuala Lumpur Internationals Conference on Biomedical Engineering, (2008).
- [5] Hiremath, P. S., Parashurambannigidda, and SaiGeeta, "Automated Identification and Classification of WhiteBlood Cells (Leukocytes) in Digital Microscopic Images", IJCA special issue on "Recent Trends in Image Processing and Pattern Recognition: RTIPPR, (2010).
- [6] Chen, Hung-Ming, Ya-Ting Tsao, and Shin-Ni Tsai, "Automatic Image Segmentation and Classification Based on Direction Texton Technique for Hemolytic Anemia in
- [7] Thin Blood Smears", Machine Vision and Applications, Vol. 25, No. 2, (2014).
- [8] Aerkewar, Prafulla N., and G. H. Agrawal, "Image Segmentation Methods for Dermatitis Disease: A Survey", Image, Vol 2, Issue 1, Pp. 01-06, (2013).
- [9] Baghli, Ismahan, "Hybrid framework Based on Evidence Theory for Blood Cell Image Segmentation", Medical Imaging, (2014).
- [10] Roussev, Vassil, and Candice Quates, "File Fragment Encoding Classification - An Empirical Approach", Digital Investigation, Vol. 10, S69-S77, (2013).
- [11] Yilmaz, Z. and M.R. Bozkurt, "Determination of Women Iron Deficiency Anemia Using Neural Networks", Journal of Medical Systems, Vol. 36, No. 5, Pp. 2941-2945, (2012).
- [12] Jolliffe, Ian, "Principal Component Analysis", John Wiley & Sons, Ltd., 2005.
- [13] WEKA Available from: <http://www.cs.waikato.ac.nz/~ml/weka>
- [14] Savkare, S. and S. Narote, "Automatic System for Classification of Erythrocytes Infected with Malaria and Identification of Parasite's Life Stage", Procedia Technology, Vol. 6, Pp. 405-410, (2012).