International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:9, No:1, 2015

Comparative Study Using Weka for Red Blood Cells Classification

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

***Abstract*—**Red blood cells (RBC) are the most common types of blood cells and are the most intensively studied in cell biology. The lack of RBCs is a condition in which the amount of hemoglobin level is lower than normal and is referred to as "anemia". Abnormalities in RBCs will affect the exchange of oxygen. This paper presents a comparative study for various techniques for classifying the RBCs as normal or abnormal (anemic) using WEKA. WEKA is an open source consists of different machine learning algorithms for data mining applications. The algorithms tested are Radial Basis Function neural network, Support vector machine, and K-Nearest Neighbors algorithm. Two sets of combined features were utilized for classification of blood cells images. The first set, exclusively consist of geometrical features, was used to identify whether the tested blood cell has a spherical shape or non-spherical cells. While the second set, consist mainly of textural features was used to recognize the types of the spherical cells. We have provided an evaluation based on applying these classification methods to our RBCs image dataset which were obtained from Serdang Hospital - Malaysia, and measuring the accuracy of test results. The best achieved classification rates are 97%, 98%, and 79% for Support vector machines, Radial Basis Function neural network, and K-Nearest Neighbors algorithm respectively.

***Keywords*—**K-Nearest Neighbors, Neural Network, Radial Basis Function, Red blood cells, Support vector machine.

I. Introduction

RED blood cells come in a variety of shapes, textures, and color depending on the types of blood disease suffered by the patient. The lack of RBCs or a condition in which the amount of hemoglobin is lower than the normal levels is defined as 'Anemia' [1]. Anemia is a common disease worldwide. WHO reports state that 66% to 80% of the world's population is suffering from anemia [2]. In Malaysia, 35% of Indian and Malay communities are afflicted with anemia [3]. In general, anemia has three main causes [4]: extreme blood loss; extreme blood cell damage, and lack of RBC production.

Two general methods can be used to identify the types of anemia [5]: kinetic and morphological.

Jameela AlKrimi is with College of Information Technology, University Tenaga National, Malaysia, and with the College of Dentistry, University of Babylon, Iraq (e-mail: T20346@uniten.edu.my).

Hamid A. Jalab is with the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia; (e-mail: hamidjalab@um.edu.my).

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

Abdul Rahim Ahmad and Azizah Suliman are with College of Information Technology, University Tenaga National, Malaysia (e-mail: abdrahim@uniten.edu.my, azizah@uniten.edu.my).

Karim Al-Jashamy is with the Faculty of Medicine, SEGi University, Malaysia (e-mail: jashamy@yahoo.com).

All blood diseases have certain effects on the shape and size of hemoglobin in RBCs [4]. Abnormalities in RBCs will affect the exchange of oxygen. Therefore, anemic RBCs can be classified based on RBCs shape abnormalities [6], [7].

Most common methods which are used for examining blood smears morphologically are expensive and be influenced the expert's expertise. For improving detection rate and increasing its accuracy, an automated algorithm for classifying the RBCs as normal or abnormal (anemic) using WEKA has been proposed.

Several attempts have been made to analysis RBCs using methods based on different machine learning algorithms. Commercially, many products (such as Cella Vision AB) have been used to count RBCs [8]. Machine Learning algorithms are used to allow computers to execute intelligent task based on databases. One of the most popular uses of machine learning in medical image analysis is classification.

Numerous WEKA applications are currently used to classify different microscopic images. WEKA is a collection of machine learning algorithms that can be used for classification and clustering. In [9] a data mining techniques was implemented with WEKA to process a medical dataset and identify the relevance of liver disorder. Four different classification methods including decision tree, Bayesian algorithms (Naive Bayes and Bayesian Networks), Neural Network classification and Rough Sets methods were used. However, the evaluation results show that using Neural Networks was the best result among the other methods.

In [10] five testing classifiers was applied using WEKA to test a set of breast cancer data. The best algorithm with an accuracy of 89.71% was Bayes network classifier compared to others.

In[11], WEKA application was used for image classification of blast cell in Leukaemia Acute Promyelocytic Leukaemia. The classification of testing images using WEKA showed that Multilayer Perceptron has 97.22% classification rate for image data collected from a Haematology unit.

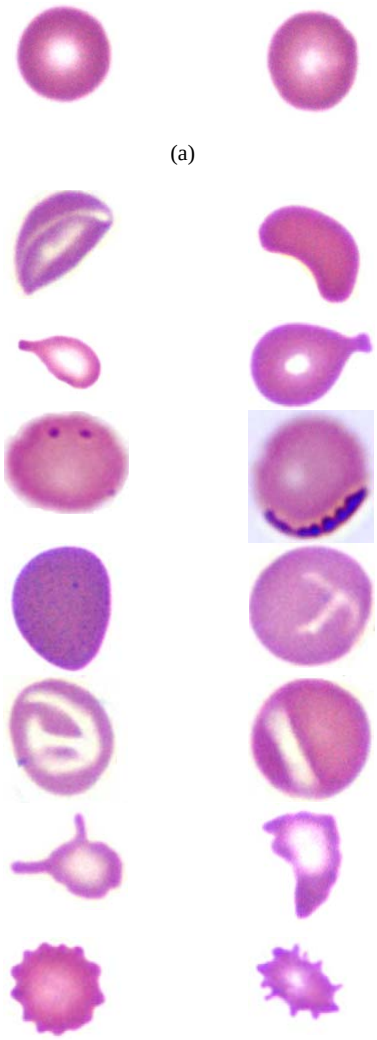
This paper proposes an automated algorithm for classifying the red blood cells as normal or abnormal (anemic) using WEKA. Total of 1000 RBCs images are processed. Geometrical and textural features are extracted for training and testing the RBFNN, SVM, and K-NN classifiers.

The paper is organized as follows: Section II describes the proposed algorithm. Sections III and IV include results and conclusions respectively.

II. Proposed Algorithm

A. Capturing of Digital Image

Peripheral blood smear slides of related anemic cases were obtained from the Hematology unit of the Pathology Department, Faculty of Medicine, Serdang Hospital from March 2012 to August 2012. All peripheral blood slides related to hematological cases were stained with May–Grünwald–Giemsa stains.



(a)

(b)

Fig. 1 Sample of RBCs images (a) normal (b) abnormal (Anemic)

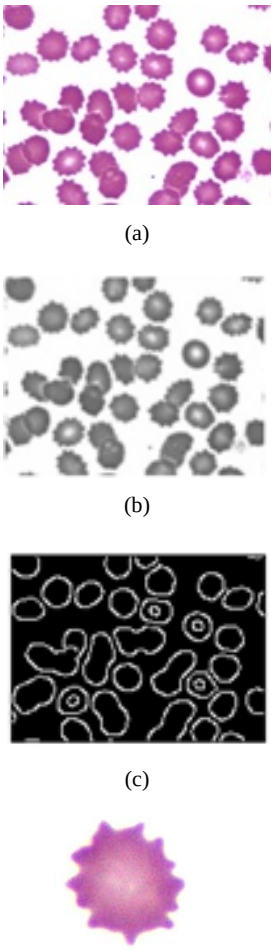
Olympus BX43 photo imaging was used convert the blood smear into a digital image through an image acquisition process. The obtained mages were manually classified into normal and abnormal (Anemic) RBCs based on the average size and shape of RBCs (Fig. 1).

Median filter was used to smoothen the images, and the k-means clustering algorithm was used to minimize the color variability of both the background and the cell pixels. Such variability was one of the problems we encountered during this work. To address this issue, we used the k-means clustering algorithm to distribute the pixel's color around two dominant colors (centroids) [12]. The images were collected (Fig. 2 (a)) and then converted to greyscale (Fig. 2 (b)). Subsequently, the binary image was obtained by using Otsu's adaptive threshold algorithm (Fig. 2 (c)). In the binary image, foreground and background image pixels were separated [13].

B. Cell Segmentation

The goal of the cell segmentation stage is to separate each RBC from the other cells in the image for further analysis (Fig. 2 (d)). In this stage, a cell in an image can be separated from the other cells with appropriate contrast from the background, and the cell's pixels are collected as one group.

Each extracted cell was cropped in four element positions using the crop rectangle, which specifies the size and position of the cropped image before the image is stored (Fig. 2 (e)).



(a)

(b)

(c)

(d)

(e)

Fig. 2 The main processing operations (a) RGB image of human blood (b) Grayscale image of human blood (c) Binary image of human blood (d) Cell image segmentation (e) The cropped cell image

C. Feature Extraction

The extracted features can be classified into two main types: The first set, exclusively consist of geometrical features, was used to identify whether the tested blood cell has a spherical shape or non-spherical cells. While the second set, consist mainly of textural features was used to recognize the types of the spherical cells. The geometrical features are the set of parameters corresponds to the boundary based and region based. Boundary based uses only the border of the RBC shape. Fourier Descriptors (FD) is most widely used in boundary based method. Texture represents the distribution of gray levels of a RBC images. The most statistical parameters representing image texture used in the literatures are: contract, variance, and moment. A total of 62 features including: one Aspect ratio; 16 Fourier Descriptors; 10 Contrast; 20 Variance, and 15 moment, as shown in Table I.

TABLE I
THE FEATURES USED TO STUDY THE CHARACTERISTICS OF RBCS

Geometrical features	Aspect ratio	1
	Fourier Descriptor	16
Texture features	Contrast	10
	Variance	20
	Moment	15

*D. Classificatio*n

The WEKA platform was used to implement our proposed algorithm [14]. WEKA is a free software tool that contains several machine learning algorithms for solving real-world problems. Three different machine learning algorithms, namely, RBFNN, SVM, and K-NN, were applied.

1. Radial Basis Function

RBFNNs are a supervised feed-forward network with one hidden layer. RBFNN has many advantages, such as rapid training procedures and improved approximation capabilities.

The RBFNN structure comprised three layers, input, hidden, and output. Source nodes were connected to the input layer, whereas the hidden layer served as a process unit for RBFNN. The weighted sum with a linear output was calculated in the output layer.

The most important task in the first stage was the selection of hidden node number. The approximation process was reduced in the presence of a small hidden node number. RBFNN has a rapid training algorithm and requires fewer training samples than other types of neural networks [10].

2. SVM Classifier

SVM is a powerful machine learning algorithm and a suitable classifier for different real-life applications including image segmentation and classification object recognition, image fusion, stereo correspondence and biomedicine. The SVM learning algorithm can enlarge the separation space between two classes to minimize the global error function [15].

3. K-Nearest-Neighbors (K-NN) Classifier

K-NN is considered as statistical learning algorithm that uses a distance to determine the nearest neighbour [16]. All training instance is considered as a point in n-dimensional space, where *n* is the number of features vector. When a test is presented, the Euclidean distance from the point represented by the test instance to each training instance is calculated. With the increasing the dimensionality of the noisy, the classification performance of KNN-distance functions is often degraded.

TABLE II
THE CLASSIFICATION RESULTS

Classifier	TP	FP	Precision	Recall	Kappa Statistic
SVM	97	0.4	98	97	0.9689
RBFF	98	0.3	98	98	0.9751
K-NN	97	0.3	97	97	0.9689

III. RESULTS AND DISCUSSION

The simulations were realized by using the WEKA. Three different machine learning algorithms, namely, RBFNN, SVM, and K-NN classifiers, were applied. To evaluate the proposed algorithm, experiments were conducted on 1000 images from our image dataset. Images collected from 200 hematology slides for different types of anemic RBCs.

All image features were verified by the faculty of medicine, SEGi University, Malaysia. The performance of the RBC classifiers was evaluated by computing the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for all tested images.

The following metrics are used to estimate the performance of the two classifiers [17]:

Precision is the percentage of RBC images that are classified as relevant.

Recall (or sensitivity) is the percentage of relevant RBC images that are labeled “relevant” by the classifier.

Out of 5000 images, 1000 were classified into two classes (normal and abnormal RBCs). All features extracted for RBC images were of 62 × 1000.

Based on Table II, we can clearly see that three classifiers showed good results, but RBFNN showed better classification rates than SVM, and K-NN classifiers. The high accuracy can be attributed to the selection of the best combination of RBC image features. An average of 970 instances out of total 1000 instances is found to be correctly classified with mean absolute error of 0.1878. Kappa statistic is used to assess the accuracy of any particular measuring cases, it is usual to differentiate between the reliability of all image features and

their validity. Based on the Kappa statistic standards, the accuracy of RBFNN, SVM, and K-NN classifiers are significant for anemic RBCs images [18]. However, using image processing and learning machine algorithms for RBCs images not only helped to decrease the cost for a clinical decision; it also enhanced the classification results.

IV. Conclusion and Future Work

In this paper, three classification methods including RBFNN, SVM, and K-NN are applied for classifying the RBCs as normal or abnormal (anemic) using WEKA. The algorithm involved the following stages: Capturing of digital image, Cells segmentation, Feature extraction, and Classification using WEKA. Compared to SVM, and K-NN classifiers, RBFNN gave higher classification results. Between the machine learning algorithms tested, RBFNN classifier can potentially and significantly improve the conventional classification methods for anemic RBCs in microscopic images. From the viewpoint of an end-user, the results of this work can facilitate laboratory work by reducing the time and cost. The future work will focus on developing a machine learning approach to classify different types of anemic RBCs in microscopic images.

References

[1] K. Parmar, M. Patel, and P. Chauhan, "A Review on Anaemia. Pharmacie Globale, 2011, 11(02), pp1-6.
[2] W. H. Organization, Global health risks: mortality and burden of disease attributable to selected major risks: World Health Organization, 2009.
[3] R. NH Nik, "The Rate and Risk Factors for Anemia among Pregnant Mothers in Jerteh Terengganu, Malaysia," Journal of Community Medicine & Health Education, 2012, Volume 2 • Issue 5,pp 1-4.
[4] S. L. Perkins, "Pediatric red cell disorders and pure red cell aplasia," Am J ClinPathol, vol. 122, pp. S70-S86, 2004.
[5] Emedicine.medscape.(2009).Available: http://emedicine.medscape.com/article/206107-overview
[6] D. Frejlichowski, "Pre-processing, extraction and recognition of binary erythrocyte shapes for computer-assisted diagnosis based on MGG images," in Computer Vision and Graphics, ed: Springer, 2010, pp. 368-375.
[7] Z. Yılmaz and M. R. Bozkurt, "Determination of Women Iron Deficiency Anemia Using Neural Networks," Journal of medical systems, vol. 36, pp. 2941-2945, 2012.
[8] A. Kratz, H.-I. Bengtsson, J. E. Casey, J. M. Keefe, G. H. Beatrice, D. Y. Grzybek, et al., "Performance Evaluation of the CellaVisionDM96 System WBC Differentials by Automated Digital Image Analysis Supported by an Artificial Neural Network," American journal of clinical pathology, vol. 124, pp. 770-781, 2005.
[9] P. M. Barnaghi, V. A. Sahzabi, and A. A. Bakar, "A Comparative Study for Various Methods of Classification," International Proceedings of Computer Science & Information Technology, vol. 27, 2012.
[10] M. F. bin Othman and T. M. S. Yau, "Comparison of different classification techniques using WEKA for breast cancer," in 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006, 2007, pp. 520-523.
[11] W. Ismail, R. Hassan, A. Payne, and S. Swift, "The detection and classification of blast cell in Leukaemia Acute PromyelocyticLeukaemia (AML M3) blood using simulated annealing and neural networks," 2011.
[12] S. Lee, M. S. Cho, K. Jung, and J. H. Kim, "Scene text extraction with edge constraint and text collinearity," in Pattern Recognition (ICPR), 2010 20th International Conference on, 2010, pp. 3983-3986.
[13] S. A. Naji, R. Zainuddin, and H. A. Jalab, "Skin segmentation based on multi pixel color clustering models," Digital Signal Processing, vol. 22, pp. 933-940, 2012.
[14] WEKA at http://www.cs.waikato.ac.nz/~ml/weka.
[15] S. Savkare 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.
[16] C. Li, S. Zhang, H. Zhang, L. Pang, K. Lam, C. Hui, et al., "Using the K-Nearest Neighbor Algorithm for the Classification of Lymph Node Metastasis in Gastric Cancer," Computational and mathematical methods in medicine, vol. 2012, 2012.
[17] P. Rajendran and M. Madheswaran, "An improved brain image classification technique with mining and shape prior segmentation procedure," Journal of medical systems, vol. 36, pp. 747-764, 2012.
[18] http://www.dmi.columbia.edu/homepages/chuangj/kappa.