Indian Journal of Science and Technology, Vol 11(46), DOI: 10.17485/ijst/2018/v11i46/137674, December 2018
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

System of Automatic Classification of Leukocytes Based on Morphological Characteristics of the Nucleus Using Computer Vision

Josede Jesus Salgado Patrón*, Oscar Quiroga Garces and Johan Julian Molina Mosquera

Department of Electronic Engineering, Faculty of Engineering, Surcolombiana University, Neiva– 410001, Huila, Colombia; josesalgadop@usco.edu.co, u20131117927@usco.edu.co, julian.molina@usco.edu.co

Abstract

Objectives: To help doctors and hematologists in the Differential Blood Count process, in order to increase productivity and eliminate human errors. **Methods:** The automation of the Differential Blood Count process offers a low-cost solution, compared to high-tech medical equipment. Due to the multiple nature of these cells and the uncertainty in the hematological images, leukocyte segmentation is one of the most important stages in this process. Scrupulous segmentation obviously reduces the errors of the following stages. In this article, we present the K-means clustering algorithm in the *Hue* – Saturation - Intensity (HIS) color space to segment the cores. In addition, the performances of three classifiers, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Quadratic Discriminated Analysis (QDA) for the recognition of leukocyte types are compared. **Findings:** In the evaluation process, the technique proposed individually is applied to each of 147 blood smear images; 139 of them were segmented with precision, reaching an average precision of 94.6%. The test consists of classifying 52 leukocytes present in images obtained in the María Auxiliadora health center, during two sessions, which contains 14 lymphocytes, 12 monocytes, 8 eosinophils and 18 neutrophils previously classified by the bacteriologist. For lymphocytes, monocytes, eosinophils and neutrophils an accuracy of 98.1%, 90.4%, 92.3% and 88.5%, respectively, is achieved. **Improvement:** The application of the proposed method shows a 92.3% accuracy of the system to classify the cells.

Keywords: Discriminant Analysis (LDA), Hue – Saturation - Intensity (HIS), K-means, Leukocyte, Quadratic Discriminated Analysis (QDA), Support Vector Machine (SVM)

1. Introduction

Microscopic images of the blood smear are routinely investigated by hematologists to diagnose most blood diseases. Three basic types of cells circulate in the blood: erythrocytes (red blood cells), leukocytes (white blood cells) and platelets. The leukocytes are, in fact, a group of different cells, with different functions in the immune system. There are five types, neutrophils (40-60%), lymphocytes (20-40%), monocytes (2-8%), eosinophils (1-4%) and basophils (0.5-1%), each with its particulariti[1].

The automatic recognition of leukocytes in hematological images usually consists of four main steps: pre-processing, image segmentation, feature extraction and classification. The segmentation of blood cells and morphological analysis is one of the most important and challenging phases in blood analysis due to the complex uncertainty of cellular nature[2]. Therefore, this step is the most challenging in many applications and the improvement of core segmentation is the most widespread effort in many investigations.

In most of the hematological images analyzed for the recognition of white blood cell types, the color of the nuclei is usually violet with different intensities and saturation levels. The HSI color space is the one used to carry out the segmentation process using the K-means clus-

*Author for correspondence

tering algorithm, since it is very similar to what we call human eye recognition. This is the reason why it is better to use HSI instead of RGB because the latter model is not closely related to the human visual perception of color.

The programming language selected is Python, which is a very simple language and has a very direct syntax. In addition, it has efficient and high-level data structures and a simple but effective approach to object-oriented programming, which makes it an ideal language for rapid development of applications in various areas and on most platforms. On the other hand, its diverse libraries provide a wide range of image processing functions, from pre-processing to classification by means of SVM, LDA, and QDA. The proposed system will help doctors and hematologists in the Differential Blood Count process, in order to increase productivity and eliminate human errors. The results were validated with the help of the bacteriologist, obtaining 92.3% accuracy.

2. Methodology

The proposed framework in Figure 1, consists of three modules: 1. image of microscopic blood smears as input, 2. processing steps (pre-processing, segmentation, feature extraction and leukocyte classification), and 3. delivery of results.

The study consists of 147 leukocyte samples taken from the María Auxiliadora health center (Garson, Huila), under the supervision of the bacteriologist. These blood smears were captured with the camera of a Samsung J5 smartphone. Digital images were taken with approximately 100X magnification factor and saved in JPG format of 2560 x 1440 pixels. The detail of the data set is provided in Table1.

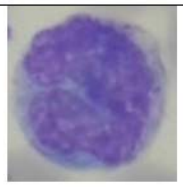
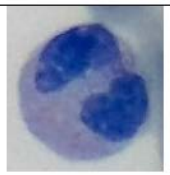
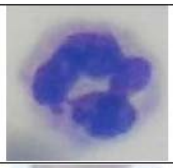
Cells	Sample
Lymphocytes	
Monocytes	
Eosinophils	
Neutrophils	
Basophils	

Figure 2. Types of Leukocytes.

The digital images were previously classified by the bacteriologist in the five classes of interest, Figure 2. The basophils were eliminated, since very few were obtained to generalize.

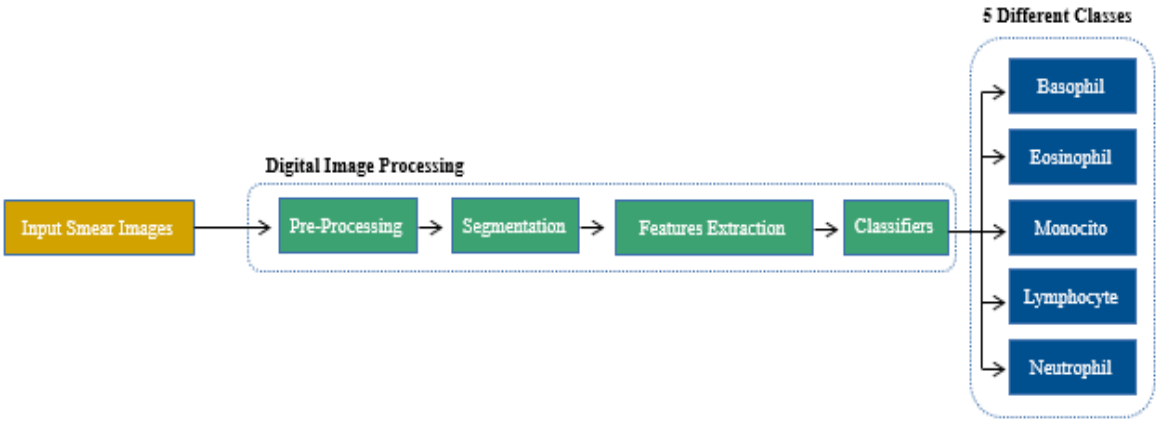


Figure1. Proposed system.

Table1. Image dataset obtained from Maria Auxiliadora Health Center

Type	No of images
Lymphocytes	51
Monocytes	15
Eosinophils	14
Neutrophils	66
Basophils	1
Total	147

2.1 Pre-Processing

In practice, color information and morphological information are used for the recognition of leukocytes[3].The different lighting can produce dissimilarities in the color of the image. However, lighting is difficult to standardize in different laboratories. So, in the first instance the matrices of the images are divided by 255 to bring the range of values between 0 and 1, and the histogram equalization is applied to redistribute the intensity of the image to cover the entire range of intensity[4]. Then, the image is converted from the RGB space to the chromatic space rg to eliminate the influence of the intensity of the illumination, Figure 3.

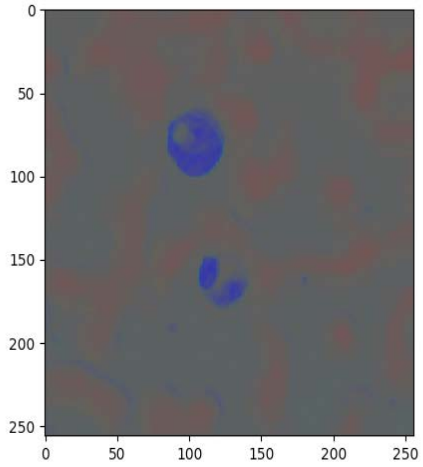


Figure 3. Rg chroma space.

The conversion of the color space of the RGB space to the chromatic space rag is described in the equation (1).

$$\begin{cases} r = \dfrac{R}{R+G+B} \\ g = \dfrac{G}{R+G+B} \\ b = 1-r-g \end{cases} \quad (1)$$

Where R, G, B are the three channel components of the original RGB image; r, g, b are the three channel components in the chromatic space rag[5].

2.2 Leukocytes Nuclei Segmentation

To segment the image, the Numpy and Scikit-Learn libraries are chosen. In the first instance, the chromatic image rg is converted to the HSI color space. The unsupervised segmentation technique is used, that is, the K-means clustering algorithm[6]. Where K specifies the number of clusters that must be created; K-means automatically assigns observations to clusters, but cannot determine the appropriate number of Clusters.

The Elbow Method is implemented to find the appropriate number of clusters[7]. This method uses the values of the inertia obtained after applying the K-means to different clusters, from 1 to N clusters, being the inertia the sum of the distances to the square of each cluster object to its centroid. Figure 4 indicates a change in K equal to 3 and represents the appropriate number of clusters to carry out the K-means clustering algorithm, Figure 5.

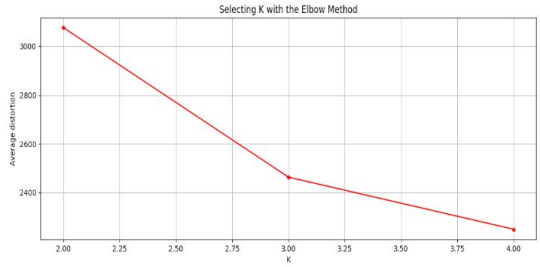


Figure 4. Elbow curve.

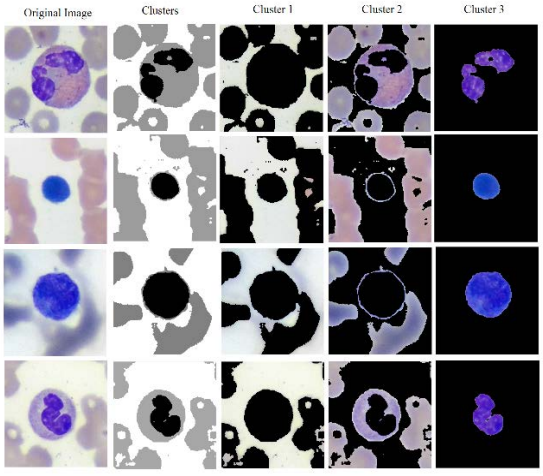


Figure 5. Appying K-means clustering algorithm on blood smear image.

Subsequently, the presence of some artifacts that are not part of the objects of interest are observed, which are eliminated with the help of a medium filter[8]. A pair of morphological filters, in effect, obtained an image with the most defined nuclei of interest. The morphological filters used are closing and opening[4], which helps strengthen weak links between objects, increase the definition of shapes and eliminate small, unwanted dark objects; also, the contours of the objects were finally smoothed.

2.3 Features Extraction

After the segmentation process, the next step is to extract the characteristics, which is a fundamental step towards the accuracy of the classification. Characteristics are descriptors of an image, which represent their natural similarities. The classifier uses these characteristics along with its labels to join different images and classify them into certain classes. The Measure module of the Scikit-image library is used to extract two different sets of features that include:

1. Geometric characteristics of an image, such as area, circularity, extension, eccentricity and number of lobes.
2. Texture characteristics such as entropy and intensity.

2.4 Classification

To classify the leukocytes to their respective subtype, characteristics describing the characteristics of the nucleus were used. We chose seven characteristics such as area, circularity, extension, eccentricity, number of lobes, entropy and intensity. In the first instance, the leukocytes were classified into two types, mononuclear or polynuclear. Characteristics such as circularity, and eccentricity, were essential to carry out this classification[9]. Subsequently, the remaining characteristics were used in the SVM, LDA, QDA classifiers to perform the classification of the leukocytes in their respective subtypes. The classifiers are evaluated using a 10-fold cross-validation. Cross-validation is a technique to evaluate how the results of a statistical analysis will be generalized to an independent data set[10]. The Fold function of the Scikit-learn library is used to validate the data. As seen in Table 2, the LDA and QDA classifiers do not differ much from each other, but show a higher percentage of accuracy compared to SVM. We chose LDA as a classifier.

Table 2. 10 fold cross validation

	SVM (%)	LDA (%)	QDA (%)
Type of nuclei	97.8	98.62	98.62
Mononuclear	98.0	98.7	98.7
Polymorphonuclear	84.0	88.0	86.0

2.5 Graphical User Interface

The application development of the automatic leukocyte classification system was carried out in the Python programming language together with the Scikit-learn and Tkinter libraries the latter aimed at designing the graphical interface for desktop applications. The graphical interface consists of a main window that is maintained during the user session, where cell detection, classification and visualization of results are performed, as shown in Figure 6.

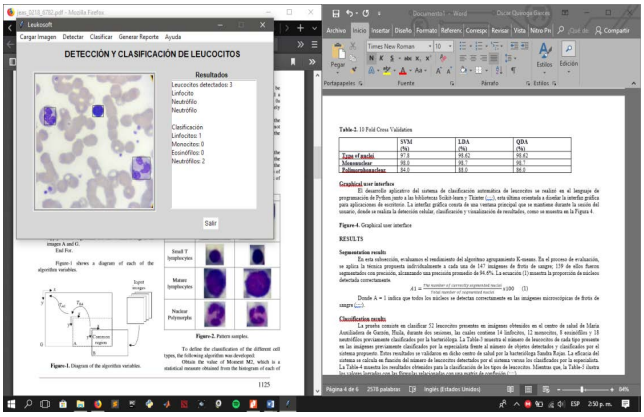


Figure 6. Graphical user interface.

3. Results

3.1 Segmentation Results

In this subsection, we evaluate the performance of the K-means clustering algorithm. In the evaluation process, the technique proposed individually is applied to each of 147 blood smear images; 139 of them were segmented with precision, reaching an average precision of 94.6%. The proportion of nuclei detected correctly is described in the equation (2).

$$AI = \frac{The\ number\ of\ currectly\ segmented\ nuclei}{Total\ number\ of\ segmented\ nuclei} \times 100 \quad (2)$$

Where A = 1 indicates that all nuclei are correctly detected in microscopic images of blood smears[11].

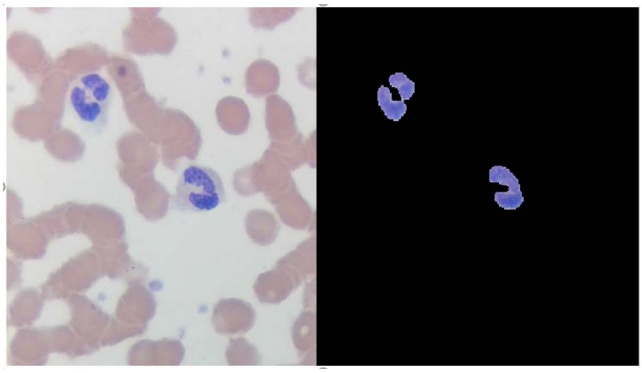


Figure 7. Segmentation.

3.2 Classification Results

The test consists of classifying 52 leukocytes present in images obtained in the María Auxiliadora health center, during two sessions, which contains 14 lymphocytes, 12 monocytes, 8 eosinophils and 18 neutrophils previously classified by the bacteriologist.

Table 3 shows the number of leukocytes of each type present in the images previously classified by the specialist against the number of objects detected and classified by the proposed system. These results were validated in said health center by the bacteriologist Sandra Rojas. The effectiveness of the system is calculated based on the number of leukocytes detected by the system versus those classified by the specialist. Table 4 shows the results obtained for the classification of leukocyte types. Table 5 illustrates the values achieved with formulas related to a confusion matrix[12].

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (3)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (4)$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

For lymphocytes, monocytes, eosinophils and neutrophils an accuracy of 98.1%, 90.4%, 92.3% and 88.5%, respectively, is achieved. With respect to the total accuracy of the proposed system, 92.3% is achieved.

Table 3. Comparison of detected leukocytes versus real data classified by a specialist

Type	Real data counting	Detected counting
Lymphocytes	14	15
Monocytes	12	15
Eosinophils	8	8
Neutrophils	18	14
Total	52	52

Table 4. Classification of the results obtained for each type of leukocyte

	TP	TN	FP	FN
Lymphocytes	14	37	1	0
Monocytes	10	37	5	0
Eosinophils	4	44	4	0
Neutrophils	12	34	2	4

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

Table 5. Values calculated for sensitivity, specificity and accuracy

Type	Sensitivity	Specificity	Accuracy
Lymphocytes	1	0.974	0.981
Monocytes	1	0.881	0.904
Eosinophils	1	0.917	0.923
Neutrophils	0.75	0.944	0.885
Total	**0.938**	**0.929**	**0.923**

4. Conclusions

It was possible to carry out the detection and classification of the cells considering only characteristics that describe the nucleus, according to their morphology. The software tool makes use of image processing and classification techniques using LDA. The automatic classification coincides in more than 92% with that made by the expert bacteriologist.It was decided to eliminate the basophils, since the number is limited according to the

images acquired in the laboratory and cannot be generalized, for which an image analysis system is presented to recognize only four groups of white blood cells in the peripheral blood.The results show that unsupervised segmentation through the K-means clustering algorithm was successfully implemented, with an accuracy of 94.6%, for more complex images, with a complicated background and containing several erythrocytes (red blood cells) close to leukocytes, likewise there are cells with multiple nuclei, where good segmentation results are obtained.In terms of classification, LDA and QDA find the hyperplane that best separates all the data points while linear kernel SVM searches for the hyperplane that best separates only the points on the border between the two classes.Because of this they were chosen to perform the classification of cells.The results are encouraging with an average of 93.8% sensitivity, 92.9% average specification and 92.3% average accuracy so that the system can be taken as a reliable, efficient and economical alternative. The latter compared to the price of the products of modern electronic systems that fulfill the same function.

5. References

1. Naranjo A, Carmen B. Atlas de Hematología Células Sanguíneas. 2nd Edicion. Universidad Católica de Manizales; 2008. p. 1–112. PMid: 18023299.
2. Shirazi S, Umar A, Naz S, Fahad SS, Razzak M. A novel method for scanning electron microscope image segmentation and its application to blood cell analysis, Journal of Applied Environmental and Biological Sciences. 2016; 6(38):90–95.
3. Zhi L, Jing L, Xiaoyan X, Hui Y, Xiaomei L, Jun C, Chengyun Z. Segmentation of white blood cells through nucleus mark watershed operations and mean shift clustering, Sensors. 2015; 15(9):22561–86. https://doi.org/10.3390/s150922561. PMid: 26370995, PMCid: PMC4610533.
4. González RC, Woods RE. Digital Image Processing. 3rd Edition Pearson/Prentice Hall; 2001. p. 1–190.
5. Nameirakpam D, Khumanthem M, Yambem JC. Image segmentation using k-means clustering algorithm and subtractive clustering algorithm, Procedia Computer Science. 2015; 54:764–71. https://doi.org/10.1016/j.procs.2015.06.090.
6. Abdul AS, Mashor MY, Rosline H. Unsupervised colour segmentation of white blood cell for acute leukemia images. IEEE International Conference on Imaging Systems and Techniques; 2011. p. 1–11.
7. Purnima B, Arvind K. EBK-Means: A clustering technique based on elbow method and k-means in WSN, International Journal of Computer Applications. 2014; 105(9):1–8.
8. Mashiat F, Jaya S. leukemia image segmentation using k-means clustering and HSI color image segmentation, International Journal of Computer Applications. 2016; 94(12):1–14.
9. Chhaya SH, Aarti GA, Samidha SK. Classification of RBC and WBC in peripheral blood smear using KNN, Paripex-Indian Journal of Research. 2012; 2(1):1–77.
10. Ramesh N, Dangott BJ, Salama M, Tasdizen T. Isolation and two-step classification of normal white blood cells in peripheral blood smears, Journal of Pathology Informatics. 2012; 3–13. https://doi.org/10.4103/2153-3539.93895.
11. Sajjad M, Siraj K, Zahoor J, Khan M, Hyeonjoon M, Jin TK, Seungmin R, Sung WB, Irfan M. Leukocytes classification and segmentation in microscopic blood smear: A resource aware healthcare service in smart cities, IEEE Access. 2016; 5:10–2940.
12. Jaroonrut P, Charnchai P. Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers, Biomed Eng Online. 2015; 14(63):1–19.