

Feature Extraction and Classification of Blood Cells Using Artificial Neural Network

Magudeeswaran Veluchamy, Karthikeyan Perumal and Thirumurugan Ponuchamy
Department of Electronics and Communication Engineering,
PSNA College of Engineering and Technology, Dindigul, Tamil nadu, India

Abstract: Problem statement: One method of evaluating the clinical status is counting of cell types based on features that it contains. There is a need for a rapid, reproducible method, superior to human inspection and for the classification of cells. For solving these problems, quantitative digital-image analysis is applied and a novel method for classifications of affected blood cells from normal in an image of a microscopic section is presented. These blood cell images are acquired from different patient with sickle cell anemia, sickle cell disease and normal volunteers. **Approach:** The segmentation of blood cells is made by morphological operations such as thresholding, erosion and dilation to preserve shape and size characteristics. These features are extracted from segmented blood cells by estimating first, second order gray level statistics and algebraic moment invariants. In addition geometrical parameters are also computed. The analysis of extracted features is made to quantify their potential discrimination capability of blood cells as normal and abnormal. The results obtained prove that these features are highly significant and can be used for classification. In addition, we use back propagation neural network to classify the blood cells more efficiently. **Results:** For testing purposes, different sizes and various types of microscopic blood cell images were used and the classification efficiency is 80% and 66.6% for normal and abnormal respectively. **Conclusion:** The proposed system has a good experimental result and can be applied to build an aiding system for pathologist.

Key words: Image segmentation, morphological operations, feature extraction, image classification, artificial neural network

INTRODUCTION

One method of evaluating clinical status is the counting of cell types based on statistical and moment invariant analysis. Cells are classified as normal and abnormal (Wheless *et al.*, 1994). Quantitative image analysis appears to offer sensitive and reproducible classification of blood cells. Quantitative digital image analysis has been previously applied to the study and classification of red blood cells (Bacus and Weens, 1977; Bacus, 1984; Gonzalez and Woods, 2009; Hasani *et al.*, 2006).

In a series of publications, Bacus and colleagues documented the utility of the technique in classifying erythrocytes from patients with various hematologic disorders. Westerman and Bacus classified cells in blood from sickle patients into 14 different classes based on six features: size (area), hemoglobin content (integrated optical density), central pallor, circularity, elongation and spicularity. This study reports the statistical and moment invariant features in the classification of cells from blood as normal and abnormal. This approach can be divided into several well-defined stages, presented in Fig. 1.

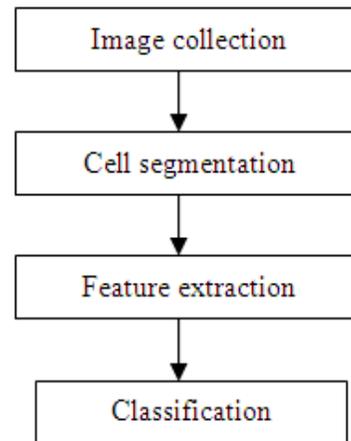


Fig. 1: General structure of method

After image collection, the image is first segmented in order to isolate the interesting parts and remove noise and undesired components. Next, the feature extraction process is applied, to extract the useful information from the segmented blood cells and

¹Corresponding Author: Magudeeswaran, V., Department of Electronics and Communication, PSNA College of Engg and Tech, Dindigul

finally the classification can be operated according to the feature extracted by the previous stage. In addition, we use back propagation neural network to classify more efficiently. In this work MBPN is used to identify the given class of input image as normal and abnormal. A number of hidden units, learning rates, momentum and Mean Square Error (MSE) are investigated to achieve the best results.

MATERIALS AND METHODS

Image Data Collection: The blood specimens were obtained from different patients with sickle cell anemia, sickle cell disease and normal volunteers. Each blood cell image contains number of normal and abnormal cells.

Blood Cell Segmentation: Image segmentation is used to detect the entire blood cells (Dougherty, 1994; Wroblewska *et al.*, 2003). In a segmented image, the picture elements are no longer the pixels, but connected set of pixels, all belonging to the same region. An object can be easily detected in an image if the object has sufficient contrast from the background. We use edge detection and basic morphology tools to detect a cell. The individual cells are close to each other and the borders among them are not well defined. The morphological operations aim at extracting relevant structures of the image by probing the image with another set of a known shape called structuring element, chosen as the result of prior knowledge concerning the geometry of the relevant and irrelevant

image structures. The most known morphological operations include erosion, dilation, opening and closing. The morphological approach to image segmentation combines regions growing and edge detection techniques (Serra, 1984; Ponsen *et al.*, 2009). The applied procedure of the image segmentation and cell separation consists of the following stages:

- Transformation of the original image into gray scale
- Detect the entire cell using edge detection technique
- Application of dilation and erosion operations to smooth the object and to eliminate the distortions

Feature Extraction: Twenty seven features were extracted from each cell image (Table 1). This included 4 geometrical features, 16 statistical features and 7 moment invariant features (Osowski *et al.*, 2004; Santinelli *et al.*, 2002).

Geometrical Features Description: We use the following geometrical features to study characteristics of the cells:

- Area A-the number of pixels on the interior of the cell
- Perimeter P-the total distance between consecutive points of the border
- Compactness C-given by the formula: $perimeter^2/area$
- Form factor $F=4*3.14*Area/Perimeter^2$

Table 1: Quantitative validation of extracted features from blood cell images

Feature	Normal cell		Abnormal cell		T-Test (P<X)
	Mean	SD	Mean	SD	
Area	618.7000000	171.5557000	896.80000	190.2070000	0.0066930
Perimeter	77.7000000	10.0005560	130.20000	25.5577780	2.93E-050
Circularity	9.9185600	0.2580060	21.16780	12.5768170	0.0057169
Form factor	1.2670800	0.0324195	0.74082	0.3331761	9.43E-050
K (No of pixels)	1050.0000000	257.4837000	2112.00000	372.2389000	9.75E-060
Mean	148.6090000	7.9401800	107.45280	26.3800700	0.0002150
Dispersion	7.4249200	1.3947620	8.37538	3.5566600	0.2309460
Variance	106.1848000	32.3851500	155.89910	137.1041000	0.1404770
Average Energy	22247.5000000	2351.0300000	12258.66000	5338.5440000	9.52E-050
Skew ness	1497.0510000	1144.5930000	-1726.39500	2998.1350000	0.0044670
Kurtosis	62969.4000000	47037.1300000	212014.50000	333606.6000000	0.0869410
Median	141.4500000	10.0151300	49.00000	67.2532500	0.0003490
Mode	147.2000000	7.1925890	107.00000	27.9195300	0.0003460
Energy	6742.6000000	4739.9486000	15647.75000	17346.4890000	0.0017242
Inertia	2474371.1000000	3383943.9000000	1570993.50000	675274.4600000	0.1219219
Entropy	346.0236100	429.3571700	144.20422	290.1908100	0.0316719
Homogeneity	162.7490000	27.0802000	394.71800	250.6820000	0.0414900
Max probability	19.0000000	13.0000000	28.00000	29.0000000	0.1000000
Inverse	266.6290000	115.5970000	381.77900	224.8940000	0.0054500
Correlation	2.8955575	2.1333436	4.43803	3.4652498	0.0187239
Φ1	1.9810100	0.0848070	1.36460	0.1260420	2.08E-080
Φ2	9.6960100	0.1253480	9.15778	0.5284510	2.08E-080
Φ3	12.8736200	0.1843860	11.97062	0.6022410	0.0003070
Φ4	7.3796900	0.1610390	5.89730	0.4597640	1.85E-070
Φ5	18.0781600	0.2611810	15.45546	0.9638120	7.48E-070
Φ6	12.7562600	0.1708900	11.05192	0.6657250	1.35E-060
Φ7	17.9084200	0.2641890	15.22908	0.9678630	6.26E-070

First order statistical features of the image: The next set of features has been generated on the basis of the intensity distribution of the image. If $u(m, n)$ is the discrete image and 'NR' is the total number of pixels in cell region. The first order gray level statistical features mean (M1), dispersion (M2), variance (M3), average energy (M4), Skew ness (M5), kurtosis (M6), median (M7) and mode (M8) are estimated. The first order gray level statistical feature formulas are given by Eq. 1-6:

$$M1(k,l) = \frac{1}{NR} \sum_{mn \in R} [u(m-k, n-l)] \quad (1)$$

$$M2(k,l) = \frac{1}{NR} \sum_{mn \in R} |u(m-k, n-l) - M1(k,l)| \quad (2)$$

$$M3(k,l) = \frac{1}{NR} \sum_{mn \in R} [u(m-k, n-l) - M1(k,l)]^2 \quad (3)$$

$$M4(k,l) = \frac{1}{NR} \sum_{mn \in R} [u(m-k, n-l)]^2 \quad (4)$$

$$M5(k,l) = \frac{1}{NR} \sum_{mn \in R} [u(m-k, n-l) - M1(k,l)]^3 \quad (5)$$

$$M6(k,l) = \frac{1}{NR} \sum_{mn \in R} [u(m-k, n-l) - M1(k,l)]^4 \quad (6)$$

If the intensity values $I(k, l)$ of $u(m, n)$ are arranged in ascending order, then Eq. 7:

$$M7(k,l) = I(k,l) \text{ of } \frac{N+1}{2} \text{ pixel} \quad (7)$$

In case of even 'N', the median value is estimated by finding the average of two middle values.

If h is the histogram of the image that gives the number of pixels with gray levels, then Eq. 8:

$$M8(k, l) = I(k, l) \text{ of } \max\{h(x_i)\} \quad (8)$$

Second Order Statistical Features of the Image:

These statistical features of second order are computed in a two step process. The first step delivers the Co-occurrence matrices containing the element. Each (i, j) th entry of the matrices represents the probability of going from pixel with gray level (i) to another with a gray level (j) under a predefined angles, 0, 30, 60, 90, 120 and 150°. Based on the co-occurrence matrices texture features are estimated. Five texture features energy (E), entropy (H), correlation (C), inertia (In) and homogeneity (L) are calculated as given by Eq. 9-13:

$$E = \frac{1}{N_\theta} \sum_{\theta} \left\langle \sum_i \sum_j P_{kl}(i, j)^2 \right\rangle \quad (9)$$

$$H = \frac{1}{N_\theta} \sum_{\theta} \left\langle -\sum_i \sum_j P_{kl}(i, j) \log[P_{kl}(i, j)] \right\rangle \quad (10)$$

$$C = \frac{1}{N_\theta} \sum_{\theta} \left\langle \frac{1}{\sigma_x \sigma_y} \sum_i \sum_j (i * j) P_{kl}(i, j) - \mu_{xx} \mu_{yy} \right\rangle \quad (11)$$

$$I_R = \frac{1}{N_\theta} \sum_{\theta} \left\langle \sum_i \sum_j (i - j)^2 P_{kl}(i, j) \right\rangle \quad (12)$$

$$L = \frac{1}{N_\theta} \sum_{\theta} \left\langle \sum_i \sum_j \frac{1}{1 + (i - j)^2} P_{kl}(i, j)^2 \right\rangle \quad (13)$$

Algebraic Moment Invariants: The classic technique for generating Invariants in terms of algebraic moments. If $u(m, n)$ is discrete input image, the algebraic moments is given by Eq. 14:

$$\mu_{pq} = \sum_m \sum_n m^p n^q u(m, n) \quad (14)$$

Here a fixed number of lower order moments for each pixel in the image are computed. Usually moment invariants are specified in terms of normalized central moments, which is given by Eq. 15:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (15)$$

Where:

$$\mu_{pq} = \sum_m \sum_n (m - \bar{m})^p (n - \bar{n})^q u(m, n)$$

$$\bar{m} = \frac{m_{10}}{m_{00}}$$

$$\bar{n} = \frac{m_{01}}{m_{00}}$$

$$\gamma = \frac{p+q}{2} + 1 \text{ for } p+q = 2, 3, \dots$$

Using above Equation for normalized central moments, a set of seven moment invariant features $\phi = [\phi_1, \phi_2, \dots, \phi_7]$ is calculated for the blood cell images.

Back Propagation Neural Network (BPNN): A back propagation neural network was created to classify the images (Markiewicz and Osowski, 2006).

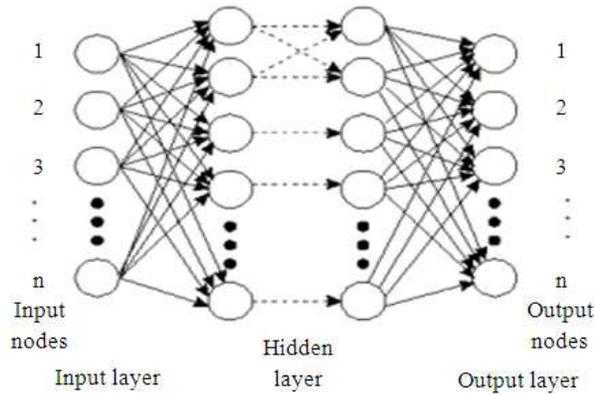


Fig. 2: Architecture of a neural network

Table 2: MBPN learning parameters and Classification rates for two classes of blood cell images

Image category	No of input neurons	No of hidden neurons	No of output neurons	Classification efficiency (%)
Normal	27	22	2	80.00
Abnormal	27	22	2	66.60

Fig. 2 shows the architecture of a neural network that contains number of input, output and hidden layers. The number of neurons in the input layer was fixed as 27, number of neurons in the hidden layer as 22 and the number of neurons in the output layer as 1. The parameters momentum and learning rate was set as 0.002 and 0.3 respectively, for training the network. The training function used was traingda. The maximum number of epochs was set as 70000. The activation function was initially set as tansig and in the hidden layer and purelin at the output layer. All these network-learning parameters are suitably optimized to achieve better classification efficiency.

RESULTS

The quantitative measures of the derived 27 features are enumerated in Table 1. The standard deviation and mean value of each feature is determined and it has been given for all blood cell classes. The recognition and classification of cells have been performed for 2 different blood cell classes. They include normal, abnormal cells. Table 2 presents MBPN learning parameters and Classification rates for two classes of blood cell images. The Fig. 3 shows blood cell image. Fig. 4 shows the pre-processed blood cell image that has well defined shape and structure. Fig. 5 presents normal blood cells. Fig. 6 presents abnormal blood cells.

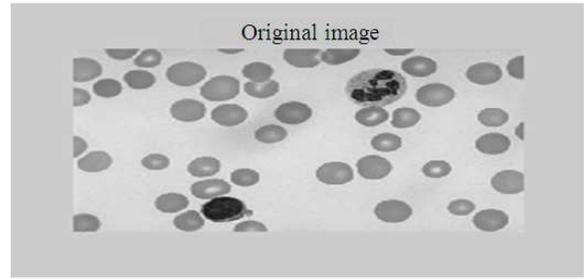


Fig. 3: Blood cell image

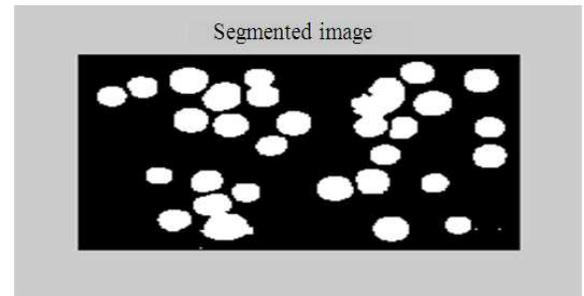


Fig. 4: Segmentation cell image

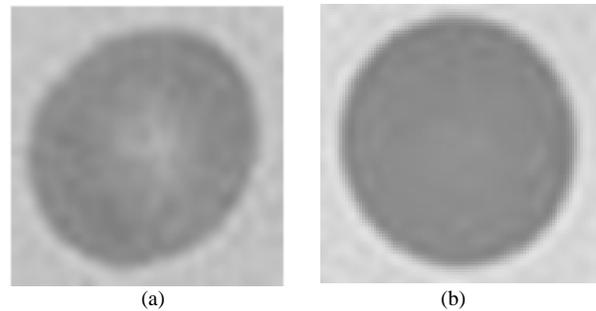


Fig. 5: Blood cells of normal

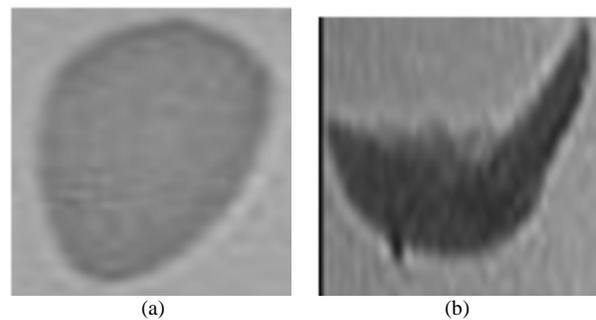


Fig. 6: Blood cells of abnormal

DISCUSSION

Our results indicate that image analysis may be used for the automatic classification of red blood cells from patients with sickle cell anemia (SS) and SC disease into the pattern classes of normal and other abnormal. The potential of feature in identifying the category is verified statistically by student t-test to evaluate 'p' value. The result attained for first order gray level statistical features shows that the features K (No of Pixels), mean, median, average energy and mode are very highly significant ($p \leq 9.75484E-06$ for K; $p \leq 0.000215$ for Mean; $p \leq 9.52231E-05$ for average energy $p \leq 0.000349$ for Median; $p \leq 0.000346$ for Mode) in discriminating the cell category which shows 95% significant difference in mean values. This followed by skew-ness is highly significant (which shows 90% significant difference) and kurtosis, dispersion and variance are less significant features (85%).

The study on features obtained by second order gray level statistical features reveals that the features energy and inverse are highly significant ($p \leq 0.0017242$, 0.00545). The remaining features of this method are less significant. In case of algebraic moment invariant features ϕ_1 , ϕ_3 , ϕ_4 , ϕ_5 , ϕ_6 , ϕ_7 perform well with a significant level of very high and remaining features of this method are less significant. In addition geometrical features perimeter, form factor are very high significant which shows 95% significant difference and remaining are highly significant.

The MPBN is trained with derived parameter irrespective of their significant level and association within the category. The network parameters, selection of number of hidden layer and hidden layer neurons are optimized to yield better classification efficiency. This provides exact classification efficiency of 80% and 66.6% for normal and abnormal respectively.

CONCLUSION

The study presents an approach for classification of blood cells as normal and abnormal in microscopic images using statistical and moment invariant analysis. Obtained results indicate that some specific features are highly significant in classification and back propagation neural network is also trained to enhance the process of classification. The network is test unknown blood cell images. The classification provided by network is appreciable. This also helps the medical technician to objectively decide on the nature of blood.

ACKNOWLEDGMENT

The researchers would like to thank all the volunteers who had participated in this study and the faculties of PSNA College of Engineering and

Technology, Dindigul for their valuable help and support to complete this study.

REFERENCES

- Bacus, J.W. and J.H. Weens, 1977. An automated method of differential red blood cell classification with application to the diagnosis of anemia. *J. Histochem. Cytochem.*, 25: 514-632. DOI: 10.1177/25.7.330716
- Bacus, J.W., 1984. Quantitative red cell morphology. *Monogr. Clin. Cytol.*, 9: 1-27. PMID: 6493236
- Dougherty, E.R., 1994. *Digital Image Processing Methods*. 1st Edn., Dekker, New York, ISBN-10: 082478927X pp: 472.
- Gonzalez, R.C., R.E. Woods, 2009. *Digital Image Processing*. 3rd Edn., Pearson Prentice Hall, India, ISBN-10: 9788131726952 pp: 954.
- Hasani, J.S.A., I.M.M.E. Emary, H.N. Heyasat and A.A. Al-zu'bi, 2006. On the application of artificial neural networks in analyzing and classifying the human chromosomes. Al- Balqa' Applied University.
- Markiewicz, T. and S. Osowski, 2006. Data mining techniques for feature selection in blood cell recognition. *Proceedings of the European Symposium on Artificial Neural Networks*, Apr. 26-28, Bruges, Belgium, pp: 407-412.
- Osowski, S., T. Markiewicz, B. Marianska and L. Moszczynski, 2004. Feature generation for the cell image recognition of myelogenous leukemia. *Warsaw University Technology*.
- Ponsen, S., N.A. Narkkong, S. Pamok and W. Aengwanich, 2009. Comparative hematological values, morphometric and morphological observation of the blood cell in capture and culture asian eel, *monopterus albus* (Zuiew). *Am. J. Anim. Vet. Sci.*, 4: 32-36. DOI: 10.3844/ajavsp.2009.32.36
- Santinelli, A., R. Mazzucchelli, P. Colanzi, A. Tinca and R. Montironi, 2002. Image processing, diagnostic information extraction and quantitative assessment in pathology. *J. Cell Mol. Med.*, 6: 93-106. DOI: 10.1111/j.1582-4934.2002.tb00314.x
- Serra, J., 1984. *Image Analysis and Mathematical Morphology*. 1st Edn., Academic, London, ISBN-10: 012637242X, pp: 610.
- Wheless, L.L., R.D. Robinson, O.P. Lapets, C. Cox and A. Rubio *et al.*, 1994. Classification of red blood cells as normal, sickle, or other abnormal, using a single image analysis feature. *Cytometry*, 17: 159-166. DOI: 10.1002/cyto.990170208
- Wroblewska, A., P. Boninski, A. Przelaskowski and M. Kazubek, 2003. Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. *Inst. Radio Elect.*, 11: 227-235.