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 (Wheeless 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](image)

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
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: \[
\frac{\text{perimeter}^2}{\text{area}}
\]
- Form factor F—\(4 \times \pi \times \text{Area}/\text{Perimeter}^2\)

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>Normal cell</th>
<th>Abnormal cell</th>
<th>T-Test (P≤X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>618.7000000</td>
<td>896.8000000</td>
<td>0.0006930</td>
</tr>
<tr>
<td>Perimeter</td>
<td>77.7000000</td>
<td>2.005560000</td>
<td>2.93E-050</td>
</tr>
<tr>
<td>Circularity</td>
<td>9.9185600</td>
<td>0.258006000</td>
<td>0.0057169</td>
</tr>
<tr>
<td>Form factor</td>
<td>1.2670800</td>
<td>0.032419500</td>
<td>9.43E-050</td>
</tr>
<tr>
<td>K (No of pixels)</td>
<td>1050.00000</td>
<td>257.4837000</td>
<td>9.75E-060</td>
</tr>
<tr>
<td>Mean</td>
<td>148.6090000</td>
<td>107.4528000</td>
<td>0.0002150</td>
</tr>
<tr>
<td>Dispersion</td>
<td>7.4249200</td>
<td>8.375380000</td>
<td>0.2309460</td>
</tr>
<tr>
<td>Variance</td>
<td>106.1848000</td>
<td>155.8991000</td>
<td>0.1404770</td>
</tr>
<tr>
<td>Average Energy</td>
<td>22247.50000</td>
<td>12258.66000</td>
<td>0.43E-050</td>
</tr>
<tr>
<td>Skewness</td>
<td>1497.0510000</td>
<td>-1726.3950000</td>
<td>0.0044670</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>62969.40000</td>
<td>212014.50000</td>
<td>0.0869410</td>
</tr>
<tr>
<td>Median</td>
<td>141.4500000</td>
<td>49.0000000</td>
<td>6.725328000</td>
</tr>
<tr>
<td>Mode</td>
<td>147.2000000</td>
<td>7.192589000</td>
<td>0.0003490</td>
</tr>
<tr>
<td>Energy</td>
<td>6742.6000000</td>
<td>15647.7500000</td>
<td>0.0017242</td>
</tr>
<tr>
<td>Inertia</td>
<td>2474371.10000</td>
<td>17346.4890000</td>
<td>0.0017242</td>
</tr>
<tr>
<td>Entropy</td>
<td>346.0236100</td>
<td>290.19081000</td>
<td>0.0316719</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>162.7490000</td>
<td>250.68200000</td>
<td>0.0419900</td>
</tr>
<tr>
<td>Max probability</td>
<td>19.0000000</td>
<td>29.00000000</td>
<td>0.1000000</td>
</tr>
<tr>
<td>Inverse correlation</td>
<td>2.8955755</td>
<td>3.438030000</td>
<td>0.0187239</td>
</tr>
<tr>
<td>Phi 1</td>
<td>1.9810100</td>
<td>1.364600000</td>
<td>2.08E-080</td>
</tr>
<tr>
<td>Phi 2</td>
<td>9.6960100</td>
<td>0.125480000</td>
<td>2.08E-080</td>
</tr>
<tr>
<td>Phi 3</td>
<td>12.8732600</td>
<td>11.97062000</td>
<td>0.0003070</td>
</tr>
<tr>
<td>Phi 4</td>
<td>7.3796900</td>
<td>5.897300000</td>
<td>1.35E-060</td>
</tr>
<tr>
<td>Phi 5</td>
<td>18.0781600</td>
<td>15.45540000</td>
<td>7.48E-070</td>
</tr>
<tr>
<td>Phi 6</td>
<td>12.7562600</td>
<td>11.05192000</td>
<td>1.35E-060</td>
</tr>
<tr>
<td>Phi 7</td>
<td>17.9084200</td>
<td>15.22900000</td>
<td>6.26E-070</td>
</tr>
</tbody>
</table>
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), skewness (M5), kurtosis (M6), median (M7) and mode (M8) are estimated. The first order gray level statistical feature formulas are given by Eq. 1-6:

\[
M_1(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} u(m,k,n,l)\]  
(1)

\[
M_2(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} [u(m,k,n,l) - M_1(k,l)]^2\]  
(2)

\[
M_3(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} [u(m,k,n,l) - M_1(k,l)]^3\]  
(3)

\[
M_4(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} [u(m,k,n,l)]^4\]  
(4)

\[
M_5(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} [u(m,k,n,l) - M_1(k,l)]^5\]  
(5)

\[
M_6(k,l) = \frac{1}{NR} \sum_{m,k} \sum_{n,l} [u(m,k,n,l) - M_1(k,l)]^6\]  
(6)

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

\[
M_7(k,l) = I(k,l) of \frac{N + 1}{2}\]  
(7)

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

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

\[
M_8(k,l) = I(k,l) of \max \{h(x)\}\]  
(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:

\[
E = \frac{1}{N_0} \sum_{i,j} \left( \sum_{i,j} P_{ij}(i,j) \right)^2\]  
(9)

\[
H = \frac{1}{N_0} \sum_{i,j} \left( \sum_{i,j} P_{ij}(i,j) \log[P_{ij}(i,j)] \right)\]  
(10)

\[
C = \frac{1}{N_0} \sum_{i,j} \left( \frac{1}{\sigma_i \sigma_j} \sum_{i,j} (i - \bar{i})(j - \bar{j}) P_{ij}(i,j) - \mu_i \mu_j \right)\]  
(11)

\[
I_0 = \frac{1}{N_0} \sum_{i,j} \left( \sum_{i,j} (i - \bar{i})^3 P_{ij}(i,j) \right)\]  
(12)

\[
L = \frac{1}{N_0} \sum_{i,j} \left( \sum_{i,j} \left[ \frac{1}{1 + (i - \bar{i})^2} \right] P_{ij}(i,j) \right)\]  
(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:

\[
m_{pq} = \sum_{m} \sum_{n} m^p n^q u(m,n)\]  
(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:

\[
\eta_{pq} = \frac{\mu_{pq}}{\mu_{oo}}\]  
(15)

Where:

\[
\mu_{pq} = \sum_{m} \sum_{n} (m - \bar{m})^p (n - \bar{n})^q u(m,n)\]

\[
\bar{m} = \frac{m_{00}}{m_{00}}\]

\[
\bar{n} = \frac{n_{00}}{m_{00}}\]

\[
\gamma = \frac{p+q}{2} + 1\]  

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

Back Propagation Neural Network (BPNN): A back propagation neural network was created to classify the images (Markiewiczl and Osowski1, 2006).
Fig. 2: Architecture of a neural network

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

<table>
<thead>
<tr>
<th>Image category</th>
<th>No of input neurons</th>
<th>No of hidden neurons</th>
<th>No of output neurons</th>
<th>Classification efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>27</td>
<td>22</td>
<td>2</td>
<td>80.00</td>
</tr>
<tr>
<td>Abnormal</td>
<td>27</td>
<td>22</td>
<td>2</td>
<td>66.60</td>
</tr>
</tbody>
</table>

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.
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 ≤ 9.75484E-06 for K; p ≤ 0.000215 for Mean; p ≤ 9.52231E-05 for average energy p ≤ 0.000349 for Median; p ≤ 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 ≤ 0.001724, 0.0054). The remaining features of this method are less significant. In case of algebraic moment invariant features φ1, φ3, φ4, φ5, φ6, φ7 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